2020

The Emergence Of Phonological Categories

Aletheia Cui
University of Pennsylvania

The Emergence Of Phonological Categories

Abstract

While phonological features are often assumed to be innate and universal (Chomsky and Halle, 1968), recent work argues for an alternative view that phonological features are emergent and acquired from linguistic input (e.g., Dresher, 2004; Mielke, 2008; Clements and Ridouane, 2011). This dissertation provides support for the emergent view of phonological features and proposes that the structure of the lexicon is the primary driving force in the emergence of phonological categories. Chapter 2 reviews the relevant developmental and theoretical literature on phonological acquisition and offers a reconsideration of the experimental findings in light of a clear distinction between phonetic and phonological knowledge. Chapter 3 presents a model of phonological category emergence in first language acquisition. In this model, the learner acquires phonological categories through creating lexically meaningful divisions in the acoustic space, and phonological categories adjust or increase in number to accommodate the representational needs of the learner's increasing vocabulary. A computational experiment was run to test the validity of this model using acoustic measurements from the Philadelphia Neighborhood Corpus as the input. To provide evidence in support of a lexically based acquisition model, Chapter 4 uses the Providence Corpus to investigate developmental patterns in phonological acquisition. This corpus study shows that lexical contrast, not frequency, contributes to the development of production accuracy on both the word and phoneme levels in 1- to 3-year-old English-learning children. Chapter 5 extends the phonological acquisition model to study the role of lexical frequency and phonetic variation in the initiation and perpetuation of sound change. The results indicate that phonological change is overwhelmingly regular and categorical with little frequency effects. Overall, this dissertation provides substantive evidence for a lexically based account of phonological category emergence.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Linguistics

First Advisor

Charles Yang

Subject Categories

Linguistics

THE EMERGENCE OF PHONOLOGICAL CATEGORIES

Aletheia Cui

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

Charles Yang, Professor of Linguistics

Graduate Group Chairperson

Eugene Buckley, Associate Professor of Linguistics

Dissertation Committee:

Jianjing Kuang, Assistant Professor of Linguistics

Don Ringe, Professor of Linguistics

THE EMERGENCE OF PHONOLOGICAL CATEGORIES

COPYRIGHT

2020

Aletheia Cui

To Tahi, the one-legged kiwi

ACKNOWLEDGMENT

The undertaking of a dissertation is so much more than a scholarly task. The work presented here would not have been possible without the intellectual and emotional support of many kind and generous people.

First and foremost, I would like to express my gratitude towards my advisor, Charles Yang, for teaching me how to approach language acquisition computationally. I came to graduate school with many vague ideas and a lot of enthusiasm about wanting to understand phonology and sound change through language acquisition, and you have helped to turn these vague ideas into concrete research questions. This dissertation very much grew out of the vowel learning project under your guidance early in my graduate career. Also, thank you for being supportive of my professional development and being understanding when I needed time to take care of myself. You are the best advisor a student could ask for.

The foundation for this dissertation comes from my fascination with sound change, and many people guided me to this point in my academic career. The start of my journey in linguistics at Cornell University was filled with many wonderful people. I would like to thank Antonia Ruppel, the world's bestest teacher of classical languages. Also, many thanks to Wayne Harbert, Michael Weiss, Carol Rosen, and Wayles Browne for allowing me to indulge in my love for ancient languages and sound change. Thank you to Sue Hertz for introducing me to speech adventures. Likewise, there have been many people that have continued to inspire me at the University of Pennsylvania. I would like to especially thank Jianjing Kuang for giving me the opportunity to study the finer details of sound change through experimental fieldwork and guiding me through my first phonetics projects. Thank you to Don Ringe for listening to my half-baked ideas about sound change and being encouraging and interested in my work, even though I have strayed quite far from historical linguistics from my undergraduate years. Thank you to Meredith Tamminga, Beatrice Santorini, and Mark Liberman for being generous with your time and data. I am also heavily indebted to the Penn Linguistics community; I have learned so much from all of you.

I would not have made it through the many trials of a Ph.D. program without my

amazing cohort: Luke Adamson, Nattanun (Pleng) Chanchaochai, Kajsa Djärv, Ava Irani, and Milena Šereikaitė. Thank you for being accepting of my introverted tendencies and yet never failing to include me. I am extremely grateful for the study sessions that got me through the first year of the program and for your continued support through qualifying papers, the dissertation proposal, and the dissertation itself. Thank you Ava, Milena, and Pleng for the many memorable trips and the unforgettable birthday parties. A special thank you to Pleng for being a fabulous cook, the most caring friend, and my last-minute dissertation editor.

My time at Penn has been made so much more colorful by a host of friends, both close and far. The biggest thank you to Emily Romanello, for teaching me Irish dance, how to make pasta, and being a very good friend. You are the one of the most diligent and nicest person I know, and your hard work and kindness will definitely pay off someday. Thank you to Jenny Proctor for nerding out with me about etymologies, Scottish Gaelic, and introducing me to music that I still listen to today. Thank you to William Diamond for tolerating my sleepiness, introducing me to coffee snobbery, and being the best rock climbing buddy. Thank you to Mike Chen for still putting up with me after so many years; I hope your blueberry bushes flourish. Also, thanks to some new friends that made my final year here less lonely: Ollie Sayeed, Gwen Hildebrandt, and Hassan Munshi. Best of luck with your pursuit of linguistics.

I am grateful to my family for letting me pursue linguistics and being supportive in spite of my unconventional career choices. Thank you always being encouraging of my academic ambitions and sowing the seeds of wanting to complete a Ph.D. when I was just starting first grade. I think I turned out all right.

Finally, I would like thank Stack Overflow for technical support. Let's be honest: This dissertation would not have gotten anywhere without all you nerds. Rock on, strangers!

ABSTRACT

THE EMERGENCE OF PHONOLOGICAL CATEGORIES

Aletheia Cui

Charles Yang

While phonological features are often assumed to be innate and universal (Chomsky and Halle, 1968), recent work argues for an alternative view that phonological features are emergent and acquired from linguistic input (e.g., Dresher, 2004; Mielke, 2008; Clements and Ridouane, 2011). This dissertation provides support for the emergent view of phonological features and proposes that the structure of the lexicon is the primary driving force in the emergence of phonological categories. Chapter 2 reviews the relevant developmental and theoretical literature on phonological acquisition and offers a reconsideration of the experimental findings in light of a clear distinction between phonetic and phonological knowledge. Chapter 3 presents a model of phonological category emergence in first language acquisition. In this model, the learner acquires phonological categories through creating lexically meaningful divisions in the acoustic space, and phonological categories adjust or increase in number to accommodate the representational needs of the learner's increasing vocabulary. A computational experiment was run to test the validity of this model using acoustic measurements from the Philadelphia Neighborhood Corpus as the input. To provide evidence in support of a lexically based acquisition model, Chapter 4 uses the Providence Corpus to investigate developmental patterns in phonological acquisition. This corpus study shows that lexical contrast, not frequency, contributes to the development of production accuracy on both the word and phoneme levels in 1- to 3-year-old English-learning children. Chapter 5 extends the phonological acquisition model to study the role of lexical frequency and phonetic variation in the initiation and perpetuation of sound change. The results indicate that phonological change is overwhelmingly regular and categorical with little frequency effects. Overall, this dissertation provides substantive evidence for a lexically based account of phonological category emergence.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The symbolic nature of phonological categories

The goal of phonology is to describe and explain the language user's abstract mental representation of speech sounds. Phonology is distinct from phonetics, which is concerned with the physical aspects of speech production and perception. Although there are different theoretical frameworks for phonological analysis, one common property that most frameworks share is the hypothesis that the *underlying representation* of speech sounds is symbolic. These discrete units of phonological representation are related to but distinct from their *surface phonetic realizations* that can be observed, recorded, and measured. One overarching goal of phonological theories is to determine what the abstract units of representation are and how they are combinatorially used in higher levels of linguistic knowledge.

The distinction between the symbolic underlying representation and the continuous signal of the surface representation is not trivial. In stating that the underlying representation for "dog" is /dɑg/, the phonologist is not merely describing that "dog" has the acoustic or articulatory characteristics of a voiced alveolar stop, a low back unrounded vowel (depending on the dialect), and a voiced velar stop. Rather, /dɑg/ is a hypothesis about the speaker's knowledge of the phonological units that make up the word "dog". The speaker knows that /d/ is a distinct speech category from /t/, or /s/, or /n/, and the speaker would recognize "tog" /tɑg/, "sog" /sɑg/, and "nog" /nɑg/ as distinct words from "dog" /dɑg/ regardless of whether these words actually exist in the English language or whether the speaker has heard these words before. Because phonological categories form a system of lexical contrast, it is

meaningless to ask whether a language learner has acquired the phoneme /t/: A phonological category does not exist in isolation but rather in opposition to other categories. The question of interest here is whether the learner has acquired some symbolic category /t/ that is distinct from other symbolic categories, like /s/, /t/, /b/, and so on.

1.2 The challenge of phonological acquisition

In first language acquisition, the signal to representation problem presents a particular set of challenges for the learner. First, the phonetic signal contains a large amount of linguistically irrelevant noise. The word "dog" /dɑg/ can have distinct acoustic realizations depending on a speaker's vocal tract length, dialect, age, their emotional state, and whether they have a cold. The same acoustic signal is also transformed by the environment it is spoken in. An utterance of "dog" /dɑg/ is different when spoken in the living room, shouted across a field, or whispered in a seminar room. Across all these conditions, the phonologically competent speaker is able to detect the linguistically relevant information from the noisy signal and retrieve the underlying representation /dɑg/ and bring into mind a friendly furry creature, likely wagging its tail. The learner, then, needs to first identify the linguistically significant cues in such noisy speech signal before they are able to create mappings between linguistically relevant acoustic information and discrete phonological categories.

While phonetic distributions of the acoustic cues can be a valuable source of information for the learner, distributional information alone cannot provide definitive separation of the categories. To illustrate this, Figure 1.1 shows vowel measurements in prosodically focused positions from child-directed speech in English (Adriaans and Swingley, 2017). A learner operating with this information alone would have trouble determining the number of categories, and it would not be obvious whether F1 and F2 are useful acoustic features at all in identifying vowels. Even though the hyperarticulated tokens in prosodically focused positions may contribute to the phonetic learning of particular tokens, the significant overlap between all the vowels suggests that distributional information is insufficient for determining the boundaries between all the vowel categories.

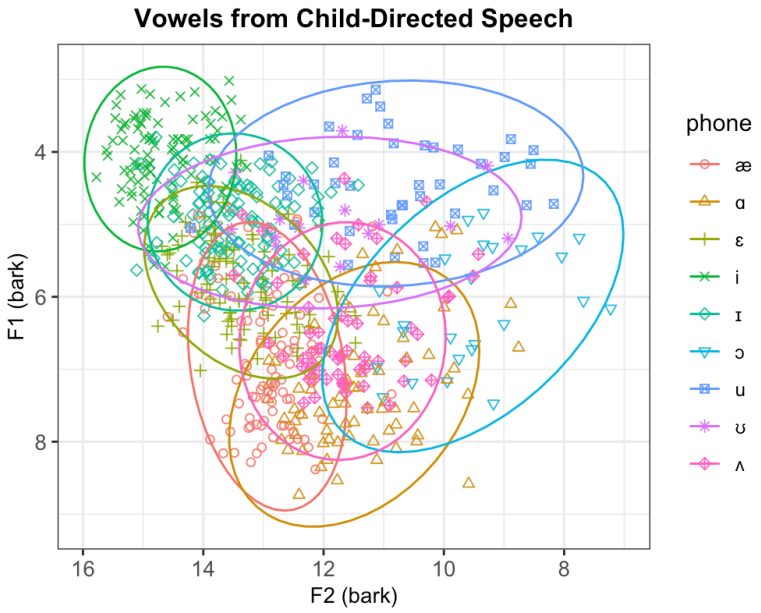


Figure 1.1: Vowel measurements in prosodically focused positions from child-directed speech (Adriaans and Swingley, 2017).

In addition to the high level of overlap in the acoustic realizations of distinct phonological categories, a single phonological contrast can be associated with a large number of acoustic cues. The English stop voicing contrast, for instance, can correspond to as many as 16 different acoustic cues in its phonetic realization. While cues such as the duration of the stop closure and the burst intensity are temporally aligned with the stop segment itself, many relevant cues for the contrast also fall on neighboring segments, such as the duration, f0, and formant transitions of the adjacent vowels (Lisker, 1986). The learner needs to not only identify that a voicing contrast exists in English but also determine which cues matter and how much they matter for each particular contrast. In addition, there is often a high degree of reduction and coarticulation in continuous speech (Johnson, 2004; Nadeu, 2014; Zsiga, 1992). In speech perception, a phonologically mature listener is able to recover reduced word forms and segments (Kemps et al., 2004; Mitterer and Ernestus, 2006) and compensate for coarticulation (Mann and Repp, 1981; Whalen, 1981; Harrington et al., 2008). However, to a phonologically naive listener, these common effects of continuous speech add an additional layer of complexity. The learner must somehow be able to construct discrete phonological

representations from very imperfect and noisy acoustic signal.

Therefore, to explain phonological category acquisition, it is necessary to look beyond distributional information in the acoustics and consider what is necessary for the learner to identify phonological contrasts and their linguistically relevant cues. The goal of this dissertation is to offer such an account of phonological category acquisition with minimal assumptions about Universal Grammar. This dissertation tests that hypothesis that phonological categories emerge from the *systematic organization* of the high-dimensional *acoustic space* to best accommodate the representation of *lexical contrast* in the learner's growing lexicon.

1.3 Outline of the dissertation

This dissertation consists of 6 chapters. Chapter 2 reviews relevant work on the nature of phonological representation and reconsiders experimental findings on first language acquisition in terms of what they reveal about the learner's phonetic and phonological knowledge. Chapter 3 presents a nonparametric and unsupervised model of phonological acquisition that learns phonological categories, their relevant acoustic cues, and lexical representations. Chapter 4 uses the Providence Corpus to provide developmental evidence in support of the proposed learning mechanism. Chapter 5 extends the acquisition model to investigate the effects of acoustics and word frequency in sound change. Finally, Chapter 6 summarizes the major findings of this dissertation and discusses possible extensions of this work.

Chapter 2

Background

Understanding the puzzle of first language acquisition has been an interdisciplinary effort. Although different approaches to studying language acquisition yield distinct insights, there is, at the same time, a lot of disconnect between these different approaches in their goals and theoretical assumptions. In this chapter, I provide an integrated discussion on a wide range of factors that matter in the acquisition of phonology and their implications for further theories and models. First, it is important to have a clear idea of what it exactly means to have acquired phonology. To do so, I consider experimental findings about phonological representations from phonetic and psycholinguistic perspectives to better understand the content of phonological categories. I argue that while phonological representations are closely related to their phonetics, they are also abstract entities distinct from their phonetic distributions. Next, I discuss theoretical approaches to phonological representation, and how different theoretical frameworks approach the acquisition problem. Then, I bring together these perspectives together to reconsider the findings from first language acquisition. Specifically, I call attention to the lack of distinction between phonetic and phonological knowledge in many developmental studies and discuss what these developmental studies tell us about the growth of phonetic and phonological knowledge. Finally, I review common conceptual and computational models to phonetic and phonological acquisition.

2.1 The nature of phonological knowledge

Over the course of first language acquisition, the learner achieves phonological proficiency, but what does it mean to have phonological proficiency? In this section, I review relevant experimental findings that shed light on the content of phonological categories in adult grammars, focusing on both the association of phonetics and phonology and the distinction between them.

2.1.1 Phonological categories are discrete units of representation

There is a correspondence between phonological units and their phonetic realizations. Acoustically, realizations of the same phoneme tend to have similar cues, and articulatorily, a phoneme is often produced with similar gestures. However, phonological knowledge about a language is not the same as phonetic knowledge, and an important distinction needs to be made between phonological categories and their physical realizations. Crucially, phonetic categories and phonological categories make different predictions about a speaker's competence: Only phonological distinctions are used contrastively in the lexicon, and these phonological distinctions are the source of a native speaker's intuition of whether certain words and speech sounds are the same or different. For example, English speakers will identify the vowels in "pit" and "pin" as the same vowel, even though the actual acoustic realizations of the vowels in these two words show significant differences. The vowel in "pin" is often nasalized in anticipation of the following nasal consonant, while the vowel in "pit" is not. Phonological units also allow speakers to normalize across acoustically distinct realizations of the same phoneme in word recognition. Allophonic variants of a phoneme tend to have distinct phonetic distributions, but they do not signal a change in meaning. A speaker can produce [bʌtɚ˞], [bʌʔə], or [bʌɾɚ˞], and these different phonetic forms map onto the same discrete lexical representation for "butter". Across phonetic and allophonic variations, there is something constant about the representation of "butter". What evidence do we have of these abstract, constant representations?

2.1.1.1 Categorical perception

First, experimental results demonstrate that there are sharp perceptual boundaries between phonemes along the relevant phonetic dimensions since listeners exhibit categorical perception along acoustic continua. The early study by Liberman et al. (1957) shows that listeners' judgment of the place of articulation would change sharply when formant transitions were synthesized in a continuum from one place of articulation to another. Although the change in phonetic realizations is gradual and incremental, listeners' responses are categorical, indicating that phonological units have shaped listeners' perception of the acoustic space. The same categorical effect in perception has been found for a wide range of contrasts, such as voicing (Liberman et al., 1961; Pisoni and Lazarus, 1974), fricatives (Repp, 1981), and tone (Peng et al., 2010). Additionally, both literate and illiterate speakers exhibit categorical perception of phonemes (Serniclaes et al., 2005). To contrast the categorical perception of phonemes, within the acoustic space of a single phoneme, discrimination between acoustically different tokens is only slightly above chance (Liberman et al., 1957). Categorical perception of continuous acoustics across phoneme boundaries indicates that listeners' perception is warped by the acoustic distribution of their language. There are several theories developed to account for this perceptual warping along phonological contrasts, such as the Native Language Magnet effect (NLM-e) (Kuhl, 1993; Kuhl et al., 2008) and the Perceptual Assimilation Model (PAM) (Best et al., 1994), which will be discussed later.

2.1.1.2 Lexical information and perceptual boundaries

Categorical perception between phonemes show that there is a strong relationship between phonemes and their associated acoustics. However, this association is not a rigid one, suggesting that the phonemes are not purely defined by their phonetic distributions alone. Listeners do not identify phonemes only based on their acoustic characteristics; they also normalize across a wide range of contextual factors. When the phonetic signal is ambiguous, listeners prefer to identify the signal with a known word. Ganong (1980) demonstrates that when presented with a VOT continuum from a word to a nonword, listeners' judgment

significantly shifts towards the word. Thus, in a continuum from *dash-tash*, listeners are more likely to choose *dash* at intermediate VOT steps. This indicates that phonological identification is closely related to word recognition, and phonological processing occurs as a interactive process, not in isolation based solely on acoustics. Additionally, perception tuning can happen based on lexical representation, further showing that phonological representation is more than knowledge of the acoustics. Norris et al. (2003) shows that when an ambiguous fricative between [f]-[s] replaces the last phone in a word with a final /f/, listeners are more likely to judge this ambiguous signal as [f] on a phoneme identification task, and the opposite is true when the ambiguous signal replaces a final /s/. Subsequent studies found that this phoneme retuning effect generalizes to novel words (McQueen et al., 2006) and voices (Bowers et al., 2016), and it happens regardless of where the ambiguous signal occurs in the word during training (Bowers et al., 2016). These results demonstrate that listeners have abstract lexical representations, and these representations can affect the interpretation of phonemic boundaries in the acoustic space. If phonological units are equivalent to their acoustic realizations, lexical factors should not have such an observable effect on phonemic identification. However, this is not the case, and phonemes exist on a level of abstraction beyond their acoustics.

2.1.1.3 Recovery of acoustically weak/absent targets

Further evidence for abstract knowledge of phonological units comes from the Phoneme Restoration Effect, where listeners recover heavily reduced or deleted phonemes. This effect was first reported by Warren (1970): When a phoneme was completely replace by a cough, listeners all perceived the missing phoneme and struggled to identify where the coughing noise occurred. Phoneme restoration is influenced by a wide range of factors. It is facilitated when the noise replace the phoneme shares acoustic characteristics of the original phoneme (Samuel, 1981). When a word undergoes extreme reduction, it is difficult to identify in isolation and additional context is needed (Ernestus et al., 2002). Although orthography can also play a role in phoneme restoration (Taft and Hambly, 1985), its influence is small

relative to the more robust phonological factors (Kemps et al., 2004). Overall, listeners' ability to recover phonemes whose acoustic cues are absent indicate that they have abstract representations of the segments in each word.

2.1.1.4 Allophones share underlying representations

Lastly, there is evidence that the underlying phonemic representation plays a role in the processing of allophones. Lahiri and Marslen-Wilson (1991) studied the processing of vowel nasalization by Bengali and English speakers. In Bengali, nasalized vowels are phonemic, while nasalized vowels are allophonic only in English. Vowel nasality led English speakers to anticipate a following nasal consonant, while Bengali speakers interpreted surface vowel nasality as signaling underlying nasal vowels. The different behaviors by Bengali and English speakers are shaped by the underlying phonology in their language. Moreover, allophonic variants can have similar effects in priming tasks. The flap [ɾ] is an allophone for /t/ and /d/ in casual speech in American English. When primed with word forms containing the flap, the processing of both the casual and carefully articulated word forms is facilitated (McLennan et al., 2003, 2005). Similarly, in the case of Spanish-Catalan bilinguals, Spanish-dominant bilinguals process [ɛ] and [e] the same because they are phonologically the same category in Spanish (Pallier et al., 2001). Overall, these experimental results show that phonological units have a close relationship with their phonetic realizations, but they are abstract entities distinct from their phonetic realizations.

2.1.1.5 The relationship between perception and production

Another important aspect of phonology is the relationship between perception and production. Gestural theories explain this relationship by positing that listeners perceive speech gestures. The Motor Theory (Liberman and Mattingly, 1985) hypothesizes a specialized module that interprets acoustic patterns into gestures, while Direct Realism (Fowler, 1986, 1996) claims that speech gestures are directly perceived from acoustic information. Part of the motivation for gestural accounts is to explain the invariant aspect of speech perception

when there is a lot of variation in the acoustic signal.

Other proposals view the the link between production and perception as abstract and phonologically mediated. Mitterer and Ernestus (2008) conducted a shadowing study with the Dutch phoneme /r/, which has two gesturally distinct phonetic variants: the alveolar trill and the uvular trill. They found that response latency did not vary if the speaker produced a different gesture for /r/ than the stimulus. In fact, the latency was longer when the speaker chose to imitate the stimulus gesture rather than using the form in their typical production. A gestural account would not be able to account for this result. Similarly, Kuang and Cui (2018a) found speech perception and production are phonologically mediated using a shadowing task with Southern Yi speakers. The vowel register contrast in Southern Yi differ both in phonation and F1. When shadowing stimuli with neutralized F1 and only distinct phonation, speakers still produced distinct F1 differences. The listener-turned-speaker must have perceived phonation differences as a cue for the register contrast, and produced the shadowed speech using their phonological knowledge of articulatory characteristics of that category. Therefore, speech production and perception must be linked by the same abstract phonological representation.

2.1.2 The content of phonological categories

Experimental evidence shows that phonological categories are more than their acoustic distributions. What, then, make up phonological knowledge? Since phonological knowledge is multifaceted, any theory of phonological acquisition should have a clear idea of the end result of this acquisition. Munson et al. (2005a) (and later reiterated in Edwards et al. (2011) outline several aspects of phonological knowledge: perceptual, articulatory, lexical/phonotactic, and sociolinguistic. A similar consideration for phonological knowledge is delineated in Pierrehumbert (2003), where the target of phonological acquisition has five levels: 1) parametric phonetics, 2) phonetic encoding, 3) lexical representations, 4) phonological grammar, and 5) morphophonological correspondences. While these are by no means exhaustive lists of what speakers actually know about their language's phonology, they are a good starting point

when studying language acquisition. In order to assess a learner's phonological knowledge, it is necessary to test their competence multiple levels. In this section, I will briefly illustrate these levels of phonological knowledge.

2.1.2.1 Perceptual knowledge

One of the most basic parts of phonological knowledge is knowing what each phonological category sounds like. This is captured by Pierrehumbert's (2003) parametric phonetics and phonetic encoding, where parametric phonetics represents the raw speech signal, and phonetic encoding provides the category interpretation of the signal. Similarly, as Munson et al. (2005a) points out, an adult English speaker knows all the acoustic cues of /s/ well enough to be able to identify the sound /s/ despite individual and environmental differences in which this sound is uttered. Additionally, this phonetic knowledge is both sophisticated and adaptable. Phonological contrasts often correspond to multiple acoustic cues (e.g., Abramson and Lisker, 1985). In determining the phonological category from the acoustics, listeners integrate multiple acoustic cues and can vary in the amount of attention they pay to each cue in perceiving a certain contrast (Beddor, 2009). Moreover, listeners often are able to compensate for coarticulatory effects to recover the intended signal, enabling stable phonetic variation, but sound change can happen when listeners shift their attention from one cue to another (Ohala, 1973; Harrington et al., 2008; Kuang and Cui, 2018b). For example, while VOT is often the most salient acoustic cue for the voicing distinction in many languages, f0 consistently co-varies with VOT. After a voiced stop, f0 tends to be lower, and after a voiceless stop, f0 tends to be higher (e.g., House and Fairbanks, 1953; Maddieson, 1984; Löfqvist et al., 1989; Dmitrieva et al., 2015). Although most speakers pay more attention to VOT in determining the voicing status of consonants (Francis et al., 2008), secondary cues such as f0 can rise in importance and eventually lead to tonogensis (Thurgood, 2002). The knowledge of cue weighting and the relative status of informative acoustic cues is an important part of perceptual knowledge.

2.1.2.2 Articulatory knowledge

In addition to acoustic characteristics, phonological knowledge also includes the articulatory gestures to make each sound. Adults know the motor sequences for producing the phonemic category of their native language across different phonetic, phonological, and prosodic contexts. Similarly, Pierrehumbert's (2003) parametric phonetic space and phonetic encoding also apply to speech production; parametric phonetic space provides the motor plan for the utterance, while phonetic encoding implements context sensitive specifics. Some theories give a more central role to speech gestures. The Motor Theory and Direct Realism propose that speech perception recovers intended articulatory gestures (Liberman and Mattingly, 1985; Fowler, 1996), and there has been experimental findings that provide evidence for aspects of the motor theory (Galantucci et al., 2006). Thus, articulatory knowledge is a crucial aspect of phonological knowledge.

2.1.2.3 Lexical representation

Another layer of phonological knowledge includes each sound's contrastive function in lexical representation (Edwards et al., 2011). The lexicon stores associations between phonological form and meaning (Pierrehumbert, 2003). With phonological categories, the speaker should be able to productively and systematically apply phonological categories in lexical representation. Speakers intuitively know when words are homophones or when words contain distinct sounds.

2.1.2.4 Phonotactics

Phonological knowledge includes an awareness of which sequences of phonemes are acceptable in their language (Munson et al., 2005a). This is termed phonological grammar in Pierrehumbert (2003). For example, English speakers know that /st-, sp-, sk-/ are possible onset clusters, but /ts-, ps-, and ks-/ are not.

2.1.2.5 Sociolinguistic knowledge

Finally, some acoustic properties of certain phonological categories can vary systematically across different social contexts. Listeners are able to use linguistic variation to infer the social identity of the speaker.

2.1.3 Phonological categories and linguistic competence

Munson et al. (2005a) and Edwards et al. (2011) argue against a strict separation between phonetics and phonology, and they view phonology as emergent through generalization across the lexicon. However, to echo points made in the previous section, the terms "phonetics" and "phonological" are necessary because they can make distinct predictions about a learner's linguistic competence. For instance, knowing that [t] and [ɾ] are distinct phonetic categories may imply that the learner knows that they have different acoustics and articulation, and they should be able to distinguish them in perceptual discrimination tasks. Having phonetic categories does not mean that the learner can apply these categories productively in distinct lexical representations. However, having /t/ and /ɾ/ as phonological categories does predict that the listener is able to use these categories in lexical representation. If the target of acquisition is phonological competency, it is crucial to mindful about the distinction between knowing phonetics and knowing phonology.

2.1.4 A note about terminology

In subsequent discussion, "phonetic category" and "phonological category" will be used as their precise definitions. When discussing previous work that is not clear about whether "phonetic" or "phonological" is intended, the term "speech category" will be used.

2.2 Theories of phonological representation

Theoretically, there are two central questions relevant for modeling phonological acquisition: 1) what is being represented and 2) what is innately available to the learner. The

nature of the representations themselves provide a guideline for diagnosing the trajectory of acquisition, and theoretical assumptions about what is innately available to the learner have important implications for models of phonological acquisition. There has been successive approaches to phonological representation of varying assumptions about the nature of phonological representation innate grammar. Earlier phonological theories posit a universal set of phonological features that are associated with articulatory or acoustic characteristics (Jakobson et al., 1951; Chomsky and Halle, 1968). More recent discussions include both more phonetically-driven approaches (e.g., Pierrehumbert, 2001; Blevins, 2004) and a sharp separation between phonetics and phonology (e.g., Hale and Reiss, 2000). These different theoretical approaches led to various proposals of how phonological acquisition occurs.

2.2.1 Innate universal features

The analysis of phonological systems of mature speakers have reveal systematic patterns and typological commonalities between different languages. The traditional generative approach view phonemes as bundles of phonological features. Jakobson et al. (1951) first introduced a system of binary phonological features in order to reduce segmental contrast into a smaller number of featural contrasts. With features, the difference between /b vs. p/, /d vs. t/, and /g vs. k/ can be described with the same distinctive feature [voice]. The subsequent influential (Chomsky and Halle, 1968) extended the feature theory and made the claim that phonological features are innate and universal. Additionally, these proposed features were intended to capture *natural classes* of sounds that pattern together in phonological processes cross-linguistically. This generative framework has been dominant in phonological research.

The acquisition framework that assumes universality was first laid out by Jakobson (1941, 1968). This framework proposes that 1) phonological acquisition occurs in a hierarchical manner, and 2) this hierarchy systematically unfolds over development in a fixed order. Jakobson (1968) appealed to both typology and common observations in child production in developing his feature hierarchy. Following this hierarchy, the first vowel a learner acquires is predicted to be maximally open /ɑ/. The first vowel contrast would be the op-

position between the low vowel /ɑ/ and the high vowel /i/, and subsequent vowels contrasts would also emerge in an orderly fashion. While Jakobson's fixed progression of phonemic oppositions provides insights into how phonological learning occurs, many studies have noted that this model cannot account for the wide range of variation in the development of individual learners (e.g., Ferguson and Farwell, 1975; Menn and Vihman, 2011). A less rigid, more phonetic proposal is that infants are born with phonetic feature detectors rather than features themselves, and these detectors allow young infants to parse acoustic signal into discrete phonetic categories (Eimas and Corbit, 1973; Eimas, 1975).

2.2.2 Issues with innate features

While assuming a system of innate features proved fruitful in phonological analysis in many aspects, there are, nevertheless, a number of issues with a strict nativist approach. One argument that has been used in favor of innate features is that very young infants show perceptual discrimination for distinctions for both native and nonnative contrasts (Eimas, 1974; Werker and Tees, 1984). This ability has been attributed to innate knowledge of phonological distinctions. However, this is not true for all contrasts. Nittrouer (2001) shows that there is a lot of individual variation in discrimination among infants and children for different native contrasts. Also, for some acoustically similar categories, language experience is needed before infants can distinguish between them. Some examples include [f] vs. [θ] and [d] and [ð] for English learning infants (Eilers et al., 1977) and [n]-[ŋ] Filipino-learning infants. Also, as Kuhl (2000) argues that this initial discriminative ability is not domain- or even species-specific. Infants can perceive nonspeech sounds categorically (Jusczyk et al., 1977, 1983), and nonhuman species also exhibit categorical perception (e.g., Kuhl, 1981; Dooling et al., 1995; Ramus et al., 2000; Mesgarani et al., 2008). Therefore, early discriminative behavior shown by young infants is likely due to general perceptual abilities rather than universal grammar.

Several other problems arise from assuming a universal set of innate features. First, the feature system was developed based on spoken languages, and it is not clear how these

features would apply to sign languages. There has been significant work on sign language features and their organization. The features proposed for sign languages are distinct from spoken language features and tend to be larger in number than spoken languages (e.g., Stokoe, 1960; Liddell and Johnson, 1989). Hierarchical organizations of sign language features has also similarly been proposed (Van der Hulst, 1993; Sandler, 1993), but these systems are also distinct from the Feature Geometry proposed for spoken languages. It is problematic that a system of universal features cannot account for both spoken and signed languages, and phonology must be language-specific and not bound to the spoken modality. Second, Mielke (2008) showed that the existing feature theories cannot adequately account for classes of sounds that pattern together in many of the world's languages, and many classes would be deemed "unnatural" based on the proposed systems of innate features. Third, there are also certain phonetic considerations that are problematic for a universal feature system. In order for a universal set of features to be able to fully account for all the possible contrasts in the world's languages, this set of features would be enormous, due to the phonetic variation between languages. There has not been a set of features that could account for all the phonological phenomena cross-linguistically (Ladefoged, 2005; Hyman, 2011). Lastly, typological universals claimed by innate feature theories often prove to have exceptions (Blevins, 2004).

2.2.3 Phonetically rich representations

One approach to tackle the issues with a universal feature set is giving phonetic realizations a central role in phonological and lexical representations (Ohala, 1990; Goldinger, 1996; Pierrehumbert, 2001). There are several motivations for proposing more phonetically driven representations. For instance, analogous phonemes in different languages have different phonetic realizations. This has been reported for Spanish vs English point vowels (Bradlow, 1995) as well as Korean vs. American English vowels (Yang, 1996) and different dialects of Portuguese (Escudero et al., 2009). Also, listeners are aware of and are affected by phonetic details in speech (Goldinger, 1998).

In exemplar models of language representation, phonological categories and lexical entries are associated with clouds of phonetic exemplars. The content of a phoneme, for example, is phonetically detailed memories of the past realizations of this phoneme. Exemplar representations are updated as new tokens are encountered. Pierrehumbert (2001) argue that generative models with a separate phonological module and phonetic implementation cannot account for variable and gradient phonetic outcomes as a result of word frequency. For example, there is no reason for t/d deletion to occur more in more frequent words in a generative model, where rules are supposedly to apply across the board. Exemplar models of speech perception and production provide a formalized way of capturing phonetic knowledge and frequency effects. It offers explanations for phonological rules that appear to be gradient or variable and accounts for frequency effects.

An exemplar-based learning model was proposed by Pierrehumbert (2003). This model is mindful of the multiple dimensions of phonological architecture, but views speech categories as probability distributions over a parametric phonetic space. Learners begins with a set number of categories, and learning occurs via the perception-production loop. If there is enough overlap between categories, these categories will merge. Pierrehumbert (2003) recognizes that a purely bottom-up approach does not account for some findings on speech perception that show top-down effects.

2.2.4 Substance-free representations

In sharp contrast with exemplar models of phonological representation, substance-free phonology steers clear of substance, i.e., phonetic details, in phonology. The fundamental idea of substance-free phonology is that phonology is a set of abstract symbols that are independent from their acoustic and articulatory properties. In this kind of framework, phonological computation occurs only over the abstract symbols of phonological categories, and physical realizations of these symbols are irrelevant in phonological processes. There is variation within the research that take the substance-free approach.

There are proposals within the substance-free framework that still strongly assume an

innate feature set (Hale and Reiss, 2003; Reiss, 2018). Hale and Reiss (2003) argues that without pre-existing phonological primitives, it would be impossible to parse linguistic input. Whereas in Jakobson's (1968) proposal, phonological features are gradually acquired, Hale and Reiss (2003) argue that learners begin with fully specified features, and unneeded features are gradually pruned. The main argument rests on the assumption of what they term to be the "innateness of primitives principle", that learning can only occur on innately available features. The child acquires phonology through collapsing contrasts that are unneeded in a process they term "lexicon optimization". The fundamental assumption that it is impossible for the learner to acquire new features cannot be added to the system is not reasonable. This proposal using innate features has the same issues as features based on phonetic substance (cf. Section 2.2.2).

Other work in substance-free phonology does not assume innate features (Odden, 2006; Blaho, 2008; Samuels, 2011). In this kind of approach, features are induced during the acquisition process. Features are acquired through phonological behavior rather than acoustic or articulatory substance. In other words, phonological features are emergent through the learning process.

2.2.5 Emergent phonological categories

There is a number of problems with innate, universal features, whether phonetically-based or substance-free. A growing body of literature that favors the idea that phonological features are emergent – that they are learned over the course of acquisition (e.g., Mielke, 2008; Clements and Ridouane, 2011). Emergent features can serve the same role innate features have in distinguishing lexical contrast and capturing common phonological processes. In innate feature theory, phonological patterns are the result of innate phonological dispositions, whereas in the emergentist approach, features arise from phonetic and phonological patterns in the language input.

Although the exact order of phonological acquisition does not follow Jakobson's proposal, the insight that phonological features are acquired in a hierarchical manner provided the

basis for subsequent development in theories of acquisition. A number of studies show that hierarchical branching trees can offer adequate description for children's developing phonology (Pye et al., 1987; Ingram, 1988b; Fikkert, 1994). Dresher (2004, 2015, 2017) built on Jakobson's idea that phonological contrasts develop in a hierarchical fashion, but he argues that features are emergent rather than innate. In Dresher's proposal, UG provides mechanisms for building up a contrastive hierarchy. He terms this process the Successive Division Algorithm ("assign contrastive features by successively dividing the inventory until every phoneme has been distinguished.") (Dresher, 2017). Rather than specifying that the first vowel as /ɑ/, it can be simply represented as /V/, a vocalic feature that can correspond to any phonetically vowel-like sounds. Further divisions of the vowel space into contrastive dimensions are language-specific, and the division can occur along any phonetic dimension, such as vowel height, frontness, or roundedness.

Along similar lines, there is more recent work arguing that linguistic contrast should play a central role in linguistic analysis and acquisition (Hall, 2007; Cowper and Hall, 2015). The acquisition of linguistic representation is a process of "assigning linguistic significance to the differences by systematically correlating differences at one level with differences at another" (Cowper and Hall, 2015). For phonological acquisition, one level of difference is phonetic, and the learner can acquire phonological categories by correlating phonetic difference with differences in word meaning. As long as the learner recognizes systematic differences on the phonetic level and can correlate it with differences in word meaning, the learner should be able to acquire phonologically relevant contrasts. The insight that linguistic contrast is of utmost importance is a rather old one (de Saussure, 1916) and needs to be pursued further in modelling phonological acquisition.

2.3 Early perceptual learning and phonetic knowledge

There is ample experimental evidence that infants are very capable phonetic learners, as shown by perceptual discrimination tasks. However, there is a general lack of discussion of how exactly perceptual discrimination is related to phonological knowledge, and success

at perceptual discrimination is often taken as equivalent as having acquired phonological distinctions. In experimental tasks that involve lexical learning, it appears that perceptual discrimination does not necessarily imply phonological knowledge, if we take the ability to use speech categories to signal lexical distinctions as an essential component of phonological knowledge.

2.3.1 Early perceptual learning

Previous research demonstrates that infants have exceptional capabilities in phonetic learning. During the first few months of life, infants are able to distinguish between most native and non-native contrasts (Eimas et al., 1971; Trehub, 1976). As early as 6 months, the perception of vowels has been shown to become more attuned towards native categories (Kuhl et al., 1992). By 10-12 months, the native language effect is clear; infants can better discriminate between phonetic distributions in their native languages and lose sensitivity to distributions not present in their native language (Werker and Tees, 1984; Polka and Werker, 1994; Best et al., 1995; Werker and Lalonde, 1988; Kuhl et al., 2006). Following these results, distributional learning was proposed as an explanation for the acquisition of phonetic categories (Maye and Gerken, 2000; Maye et al., 2002; Vallabha et al., 2007; Maye et al., 2008; McMurray et al., 2009; Werker et al., 2012). Maye et al. (2002) showed that 6- and 8-month-old infants could learn to discriminate acoustic tokens when they were from a bimodal distribution but not a unimodal distribution. These results suggest that the natural clustering of acoustic cues can play a role in early perceptual reorganization in the first year of life. However, since their experiments are phonetic in nature, it is impossible to conclude whether these results directly transfer to phonological knowledge. Moreover, Cristia et al. (2011) shows that there are limitations for distributional learning; although 4- to 6-month-old infants succeeded in learning the retroflex place of articulation, their failed to learn alveolo-palatal place of articulation using distributional acoustics.

2.3.2 Phonetic awareness and lexical learning

Experiments on early word learning show seemingly conflicting results. On the one hand, it is clear that infants are aware of phonetic details in their early representation of lexical items. Young children's phonetic knowledge of words is often tested via mispronunciation tasks, where the pronunciation of a familiar word is modified by phonologically contrastive features. As young as 11 months, infants were able to detect mispronunciations in familiar words (Swingley, 2005). Similar results have been found for a range of ages early in language acquisition (Bailey and Plunkett, 2002; Swingley and Aslin, 2002; Swingley, 2009; Mani and Plunkett, 2010). So far, these findings suggest that young children apply their phonetic competence in word learning, and their discrimination skills may have contributed to the detailed phonetic knowledge of words.

On the other hand, young learners sometimes struggle to learn phonologically similar words. One commonly cited study that demonstrated a failure in learning similar-sounding words was Stager and Werker (1997). In this study, Stager and Werker (1997) used a switch task to study whether infants are able to learn minimal pairs. The infants are first familiarized with sound-object pairings; during the test portion, the sound-object pairings are changed. If the infant showed a longer looking time, it indicates that they noticed the switch. Using this procedure, 14-month-olds were able to distinguish words that differed by multiple phonemes (*lif* vs. *neem*) but failed to notice the switch when the words differed by one phoneme ([bɪ] and [dɪ]). A number of other studies found similar results (Werker et al., 2002; Pater et al., 2004; Nazzi, 2005). Interestingly, the same 14-month-olds were able to distinguish between [b] and [d] in a phonetic discrimination task. Also at this age, when tested with familiar words like *ball* and *doll*, 14-month-olds were able to detect when a word did not match the object (Fennell and Werker, 2003).

Experimental results with adult speakers confirm that phonetic discrimination is very different from have categorical phonological distinction. It is well-known that Japanese speakers often fail to distinguish between [r]-[l]. Nevertheless, they are capable of perceiving equivalent F3 cues in nonspeech stimuli (Miyawaki et al., 1975). Similarly, English speakers

do not have phonemic click contrasts, and yet they are able to discriminate Zulu clicks with good accuracy (Best et al., 1988). The ability to notice acoustic differences between sounds is very different from the linguistic processing of acoustic information.

From the results summarized so far, there are two major themes. First, young language learners have very good phonetic discrimination skills and detailed phonetic knowledge of familiar words. Second, despite the good phonetic knowledge, they struggle to learn new words that are phonologically similar. If they had indeed acquired phonological distinctions early on, why are new phonologically similar words difficult to learn? To resolve these conflicting results, it is necessary to consider these findings by clearly distinguishing between what is learned phonetically and phonologically.

2.3.3 Phonetic vs. phonological categories

The imprecise use of linguistic terminology is very common in developmental studies. For example, Maye and Gerken (2000) presents a distribution-based approach for "phonemic learning" and only to later note that "[t]he categories we would like to account for are not 'phonemes,'...What we are interested in is perhaps more appropriately termed 'phonetic categories' or 'phonetic equivalence classes'". Maye et al. (2008) refers to the experimental conditions in Maye et al. (2002) as "single phoneme vs. two alternating phonemes", while the experimental design and language in Maye et al. (2002) was entirely focused on phonetic discrimination and made no claim about phonological learning. In computational models of acquisition, there is similarly seldom discussion about the difference between phonetic and phonological learning. McMurray et al. (2009) only mentions phonemic categories at the very beginning, and then talks about "phonetic category" and "speech category" learning. Feldman et al. (2013a) presents a model of "phonetic category acquisition" and yet frequently refers to "phonemes" in their discussion. There is only explicit discussion about phonetic and phonological categories in Dillon et al. (2013), who argued that the acquisition literature implicitly assumes phonetic and phonological category acquisition as a two-stage process.

How important is it to stress the difference between phonetic and phonological cate-

gories? After all, phonemes are typically associated with some phonetic distribution. However, the difference between "phonetic category" and "phonological category" is not a trivial one as they make distinct predictions about what the child can do with the said category. As discussed in earlier sections, one of the defining characteristics of a phonological category is its ability to signal lexical contrast. Exchange the /p/ in "parrot" /pæɹət/ for /k/, any English speaker would immediately recognize that "carrot" /kæɹət/ refers to something very different that "parrot". Thus, claiming that a child has learned a phonological category predicts that the child will be able to use this category *productively* (i.e., in representing new words) and *systematically* (i.e., able to apply the same representation to all words containing this category). However, knowing phonetic categories does not make these predictions because phonetic categories do not indicate lexical contrast.

If perceptual discrimination were taken as evidence for the acquisition of a phonological contrast, 14-month-olds should not have failed to identify [bɪ] and [dɪ] as distinct words. Even though toddlers know *ball* and *doll* are different, it appears that their knowledge is specific to these words. In other words, they have not acquired /b/ and /d/ as distinct categories since they fail to generalize the difference to novel words. If we do not equate perceptual discrimination with phonological knowledge, the results that learners sometimes fail to use discriminable phonetic details in word learning is not only unsurprising, but expected.

An analogy can be made with visual perception. When presented with two slightly different shades of blue, the subject is likely to respond that the two colors are different. However, if the subject is asked to identify the color for both shades, the answer is likely "blue" for both. MacKain (1982) offers some early criticism of the use of discrimination paradigms to draw conclusions about language acquisition. Indeed, phonological developments continue far past the first few months and even years of language learning. Experiments with older children show that they differ from adult level competence even by age 12 (Hazan and Barrett, 2000).

2.4 Factors in phonological development

As discussed in Section 2.1, phonological knowledge is abstract and multifaceted. With this in mind, in this section, I discuss the various factors that contribute to the mapping between acoustic forms and word meaning, and how this relates to the acquisition of phonology.

2.4.1 Acoustic salience

Acoustic salience plays a role in perceptual learning and word learning. Although infants initially are able to discriminate between most native and non-native contrasts, some sounds are inherently more psychoacoustically similar. More language experience is often required before a child is able to discriminate between these sounds. For instance, English learning 1- to 3-month-olds and 6- to 8-month-olds struggled to distinguish between the acoustically similar [fa] vs. [θa] and [fi] vs. [θi] (Eilers et al., 1977). Polka et al. (2001) showed that even at 10 to 12 months, English learning infants were unable to consistently discriminate between [d] and [ð]. In a study of nasal place of articulation, Narayan et al. (2010) found that native language experience was required for Filipino-learning infants to discriminate [na]-[ŋa], but not the acoustically more distinct [ma]-[na]. Although infants generally have good phonetic learning skills, not all phonetic distributions are learned at the same pace. Even though learners of the above-mentioned languages all succeed in acquiring their native phonology, the difficulty in perceptually discriminating between certain sounds may delay the acquisition of those sounds.

The degree of acoustic salience between similar-sounding words can facilitate the mapping of acoustic forms to their referents. Curtin et al. (2009) showed that 15-month-old infants learning Canadian English were able to learn minimal pairs that had the vowel contrast /i/-/ɪ/, but not /i/-/u/ or /ɪ/-/u/. The vowel pair /i/-/ɪ/ differ in F1, while the most contrastive acoustic dimension for /i/-/u/ and /ɪ/-/u/ is F2. Curtin et al. (2009) suggest that infants at 15 months pay attention to the more salient (i.e., F1, which is lower in the frequency spectrum) cues when learning new words. A similar study in Escudero et al.

(2014) provided further evidence that acoustic salience plays a role in word learning. In this study, 15-month-old Australian English learning infants were presented with non-word stimuli with the vowels /i/, /ɪ/, and /u/ in either Australian English or Canadian English. The acoustic distinctions between these three vowels are greater in Canadian English. The Canadian English condition infants were able to detect that words with /ɪ/ and /u/ as distinct from /i/, while the Australian English condition infants could not. These results suggest that the acoustic salience of phonological contrasts play a role in the acquisition process.

2.4.2 Top-down information

Top-down lexical, syntactic, and contextual cues can also contribute to the mapping of acoustics to phonological forms. Although children failed to learn similar-sounding words in certain experimental conditions (e.g., Stager and Werker, 1997; Werker et al., 2002; Pater et al., 2004), other studies show that similar-sounding words can be learned from sound-object pairings. Yoshida et al. (2009) found that when tested with a visual choice paradigm rather than the switch task in Stager and Werker (1997), 14-month-olds were able to learn "bin" and "din" as separate words. In another study, Yeung and Werker (2009) conducted three experiments to investigate the effect sound-object pairings in learning speech categories. In a perceptual discrimination task, they demonstrated that 9-month-old English learning infants could no longer distinguish the Hindi dental stop [d̪a] and retroflex stop [ɖ̪]. Then, in a second experiment, infants were presented with the tokens with this contrast consistently paired with distinct visual stimuli. After familiarization, infants gained the ability to discriminate between the two nonnative sounds. A third experiment was conducted in a similar set up, but the sound-object pairing was inconsistent. In this experiment, infants failed to learn the difference between these two sound categories. While much of the previous work suggested statistical learning as the primary source of perceptual reorganization, Yeung and Werker (2009) show that the association between sound and meaning is also an important part of this process. Although distributional cues exist in the stimuli, their

experiment 3 provides evidence that distributional cues alone cannot trigger the learning of a contrast.

Moreover, young children are more likely to succeed in learning similar-sounding words when with clear cues on other linguistic levels. Fennell and Waxman (2010) showed that 14-month-olds were able to learn "bin" and "din" as distinct words when the referential contexts were clear. Syntactic and semantic information can also contribute to the identification of contrast. Dautriche et al. (2015) demonstrated that French 18-month-olds are able to more easily learn new nouns that are close in phonological form to a familiar verb rather than a familiar noun. The semantic and syntactic differences, coupled with phonetic differences, helped these subjects conclude that the novel phonetic form was indeed a new word. Peperkamp and Dupoux (2007) conducted an artificial language learning experiment to study the acquisition of phonemes with allophonic variation. The results indicate that the subjects learned allophones as a single underlying phoneme when exposed to the same semantic information, and they were able to generalize the learned phoneme to novel lexical items. However, when only acoustic cues were present, they were not able to generalize the learned category to novel lexical items. It appears that phonological learning only occurred with semantic information, and phonological generalization could happen even when the allophonic groupings were "unnatural".

Altogether, these results indicate that the learning of similar-sounding words is possible if the child had access to clear top-down lexical, syntactic, or contextual information.

2.4.3 Vocabulary growth

Recent work has shown that word learning begins at an early age, around the same time perceptual tuning occurs. As early as 4.5 months old, infants already show preferences for their own names (Mandel et al., 1995), and 6-month-olds look at appropriate figure in video upon hearing *mommy* or *daddy* (Tincoff and Jusczyk, 1999). Also at 6 months, infants can use familiar words like their own names and "mommy" in word segmentation (Bortfeld et al., 2005), and they can segment words that occur at utterance boundaries (Johnson

et al., 2014). At 6-9 months, infants know the meaning of some common words (Bergelson and Swingley, 2012), and 7.5-month-olds are able to detect common words in fluent speech (Jusczyk and Aslin, 1995). Additionally, 8-month-olds have been shown to remember words two weeks after exposure (Jusczyk and Hohne, 1997).

Children with larger vocabularies find it easier to learn new words. Several studies discussed previously included measures of vocabulary, and they found that children with larger vocabularies are more likely to success at learning phonologically similar words (Werker et al., 2002; Yoshida et al., 2009; Mani and Plunkett, 2010). For 16-month-olds, the ability to learn novel words is correlated with their expressive vocabulary size (Horváth et al., 2015). At 2 years old, children with larger vocabularies are more likely to treat a word that is phonologically similar to a familiar word as a novel word (Swingley, 2016), and a similar effect was found for slightly older children between 30-46 months (Law and Edwards, 2015). It is possible that some children have larger vocabularies simply because they have a tendency to confer new word status to novel acoustic stimuli. Another possibility is that they are more advanced phonologically, enabling them to generalize learned phonological distinctions to more easily recognize and represent new words. However, these two options are not mutually exclusive. A child that is more likely to associate new acoustic forms with new words would have more items in their lexicon to make phonological generalizations from, and with more advanced phonology, the learner can better identify whether a novel acoustic form contains a meaningful difference from the words they already know.

Having a larger vocabulary can also contribute to better word processing. Word recognition is faster and more accurate for 18- and 21-month-olds with larger expressive vocabularies (Fernald et al., 2001). In a longitudinal study of children at 15, 18, 21, and 25 months, Fernald et al. (2006) found that the growth of expressive vocabulary in the second year of life enhances word recognition at 25 months. In a follow-up study when the subjects were 8 years of age, the vocabulary size and word recognition results at 25 months predicted the variance in their linguistic and cognitive skills (Marchman and Fernald, 2008). Studies of early and late talkers also show that vocabulary size and phonological ability go hand-in-

hand. When compared with their peers at the same age, late talkers have both smaller vocabulary and less advanced phonological systems (Stoel-Gammon, 1991; Paul and Jennings, 1992; Rescorla and Ratner, 1996). On the other hand, precocious talkers have both larger production inventories and better production accuracy (Smith et al., 2006).

It is evident that vocabulary development and phonological acquisition are closely related processes. The question remains as to how the child forms phonological generalizations from learning words and their acoustics.

2.4.4 Early lexical representation and phonological generalization

There is some experimental evidence for the development of abstract representations. Early on, infants fail to generalize across cues such as speaker voice and variation in pitch. At 7.5 months, English-learning infants only succeeded in recognizing familiarized words if the voice of a new speaker is similar to the one they were familiarized with; when the sex of the speaker was changed, they failed to recognize the familiarized words (Houston and Jusczyk, 2000). Also at this age, English-learning infants are able to recognize the same word presented with different amplitudes but not when the pitch was varied (Singh et al., 2008). It appears with limited language experience at 7.5 months, infants have difficulty identifying acoustic information that is stable for lexical identity. Slightly older infants were able to generalize across contextual and indexical cues. For example, English-learning infants are able to generalize across pitch differences at 9 months (Singh et al., 2008) as well as across gender and affect at 10.5 months (Houston and Jusczyk, 2000; Singh et al., 2004). These results show that infants gradually learn to distinguish between cues that are informative for lexical identity and other cues.

What is the nature of early lexical representation? One view is that children initially learn whole-word patterns, with phonological categories emerging from the network of known words (Vihman and Keren-Portnoy, 2013). In child production, the same word and phoneme can be produced with a large range of variation, although typically retaining some of the features from the adult model (Ferguson and Farwell, 1975). This indicates that the child

may not treat the sounds in a word as a sequence of phonemes at this early stage, and their early representations are more likely holistic impressions of salient acoustic features over some larger unit. In a longitudinal production study of four children aged 1-2, Sosa and Stoel-Gammon (2006) also found that there was a large amount of variability in the production of words, and there was no general decline in the amount of variability, but rather variability occurs in peaks and valleys. The increase in variability has been interpreted as the manifestation of emerging systematicity in phonology (Vihman, 2014).

It is interesting that while perceptual studies show that young learners have detailed phonetic representations of words, production results suggest that the representation is a rough impression of the sounds in a word. Since speech production and perception is mediated by phonology, the likely explanation is that their phonology is only at the initial stages of development. As a result, they can have detailed phonetic knowledge of words, while at the same time, their inability to consistently interpret the phonetics of a word in phonological units results in inconsistent production. The question remains as to the exact nature of early phonological representation and how it is related to phonetics and lexical knowledge.

2.5 Previous models of speech category acquisition

Similar to the lack of integration and consideration for phonetic and phonological knowledge in developmental work, conceptual and computational models of acquisition also tend to follow this trend and fall into two campus. There is one line of work that mainly investigates the acquisition of categories on the phonetic level. On the other hand, there are models that aim to explain higher phonological structures, typically within some linguistic framework. The approaches and assumptions are generally very different for these two kinds of modeling.

2.5.1 Phonetic models of acquisition

There is a number of conceptual frameworks and computational models that address the formation of phonetic categories. These models explain the perceptual tuning that have been

observed within the first year of life where the perception of native categories is enhanced while the perception of nonnative categories is diminished (e.g., Eimas et al., 1971; Trehub, 1976; Werker and Tees, 1984).

The perceptual assimilation model (PAM) developed by Best and colleagues addresses the decline of non-native contrast perception through articulatory phonology (Best et al., 1988; Best, 1993). Specifically, the perception of non-native categories declines when their articulatory gestures can be interpreted as similar gestures to a native categories. This model predicts that non-native contrasts that have distinct articulations from native ones should be easier to perceive than non-native categories that can be assimilated. On the other hand, the Native Language Magnet model (NLM) focuses on tuning towards native categories in perception rather than the loss of non-native categories (Kuhl, 1993). In this model, infants are first able to distinguish between native and non-native categories by using their general auditory processing abilities. Subsequently, with language experience, infants learn from the acoustic distributions and thus form perceptual magnets, i.e., their perception becomes warped towards native categories. Although both PAM and NLM can explain the changes in the perceptual abilities of infants, both models are phonetic in nature and only explain part of the phonological acquisition puzzle. Neither model addresses how changes in perception translates into abstract units of phonological representation.

In addition to conceptual models, there has been a number of computational models that address language acquisition at the phonetic level. Most of them do not sufficiently address the relationship between signal and abstract phonological acquisition. Many such computational models rely on statistical learning. These studies are, for the most part, also motivated by findings from perceptual tuning within the first year, where distributions of acoustic information contributed to the ability to perceptually discriminate between categories in the ambient language (e.g., Werker and Tees, 1984; Kuhl et al., 1992; Polka and Werker, 1994; Kuhl et al., 2006). The models that adopt distributional learning as the learning mechanism often focus on acoustic information as the primary cue of category learning and treat category acquisition as a clustering problem.

One common approach is to treat sound categories as a mixture of Gaussians. De Boer and Kuhl (2003) used the standard expectation maximization (EM) algorithm (Demuth et al., 2006) to learn the English vowel categories /i, ɑ, u/ from infant-directed speech and adult-direct speech. They found that infant-directed speech produced better learned clusters, but this should be taken as an acquisition model since the number of categories were pre-specified in the model. Vallabha et al. (2007) presented two algorithms to learn a subset of English and Japanese vowels. The Parametric Algorithm for Online Mixture Estimation (OME) is an online version of EM that relies on the assumption that vowel acoustics are drawn from multivariate Gaussian distributions. The second algorithm, Topographic OME (TOME), does not rely on vowels acoustics as Gaussian distributions. Lake et al. (2009) applied OME to a number of category learning tasks and argue that OME can works as an acquisition model for human category learning. Adriaans and Swingley (2017) implemented vowel learning as the discovery of multivariate Gaussian categories, with parameter estimation using EM. Two sets of vowels were learned: the point vowels /i ɑ u/, vs. close vowels /i ɪ ɛ/. There were three conditions for each: baseline on all tokens, prosodic focus set, and no prosodic focus set. Unsurprisingly, the focus condition had better accuracy than all tokens and no focus, and point vowels were learned better than close vowels. Also using a mixture of Gaussians approach, McMurray et al. (2009) introduced a competition mechanism to account for the pruning and enhancement of category learning. In a similar manner, Toscano and McMurray (2010) modeled cue integration for phonological categories for the voicing contrast.

While these model succeeded in discovering categories from the acoustic input, there is a general lack of clarity in whether the end result of learning is phonetic or phonological. There are very few studies that explicitly discuss the distinction between phonetic and phonemic categories. In one of such study, Dillon et al. (2013) points out that statistical learning approaches discover phonetic, rather than phonemic categories. They argue that such statistical approach implicitly suggests a two-stage process in the acquisition of phonology where phonetic categories are learned before phonological ones.

2.5.2 Phonological and integrative models

There is a number of linguistically motivated models of phonological acquisition. Unlike phonetic learning models that often make use of distributional acoustic information alone, many of the phonological models use discrete phone or feature level representations as the input. For example, Peperkamp et al. (2006) implemented a statistical learning model of allophones based on complementary distributions. Their experiments used phone and feature level transcriptions as input. Similarly, the simulations in Boersma and Hayes (2001) also used discrete phone level transcriptions in an Optimality Theory based learning model.

Although the goal of theoretical linguistics is to explain language learnability by mapping out the hypothesis space of grammar, there has been relatively less work directly studying language acquisition within the field of linguistics. The early work by Jakobson (Jakobson, 1941, 1968) on phonological acquisition using innate phonological features has been discussed in detail Section 2.2.1. More recent work on phonological acquisition used the Optimality Theory framework. For example, Boersma et al. (2003) described an account of perceptual warping and abstract phonological category learning using OT. To account for perceptual tuning within the first year of life, they posited the constraint *Warp that allows certain regions of acoustic space to be perceived as more similar. Next, lexical items guide the transformation of these phonetic categories onto discrete phonological features. While not a learning model, Hayes (2004) interprets the results of experimental work on language acquisition within the OT framework. The perceptual tuning is interpreted as the acquisition of the ranking of the constraint Ident, and early sensitivity of phonotactics is interpreted as evidence for the acquisition of markedness constraints with respect to Max and Dep.

There are more integrative models that have been developed to account for the relationship between phonetics and higher levels of phonological representation. Jusczyk's (1997) WRAPSA (Word Recognition and Phonetic Structure Acquisition) is conceptual model that bridges phonetics and word learning. In this model, the infant starts with a set of global auditory analyzers that process both speech and non-speech auditory signals. Next, a weighting scheme is determined based on distributions and features of the sounds.

Given the weighting scheme, the learner can extract prosodic units from continuous speech. Werker and Curtin's (2005) PRIMIR (Processing Rich Information from Multi-dimensional Interactive Representations) posits three levels of representation, or planes, that interact in learning: General Perceptual (all the information from the signal), Word Form (exemplar-based representations of phonetic forms without meaning attached), and Phoneme (emerges from generalizations from clusters on the Word Form plane). In PRIMIR, the Phonemic plane emerges as the learner gains experience with word forms that share similar phonetic features. The learner is equipped with three dynamic filters, which are initial biases, the learner's developmental level, and requirements of the specific language task. More recently, Vihman (2017) describes the complementary systems model drawn largely from research in brain development and memory function. This model divides phonological acquisition into implicit, or distributional learning, and explicit, or declarative learning. The implicit learning is not a conscious process and advances phonological development by drawing generalizations based on the distributional cues, while explicit learning, which requires conscious focus, enables the learner to recognize the form and meaning of words. Crucially, these processes occur in parallel.

On the computational side, there has also been work integrating phonetic and phonological category learning with other levels of acquisition. Using Bayesian models, Feldman et al. (2013b) simulated the simultaneous learning of vowels and words, and Elsner et al. (2013) investigated the acquisition of word segmentation, lexical items, and phonetics. There has also been some work on how features emerge in language acquisition. Lin (2005) used a mixture of hidden Markov models to learn features, segments, and words from waveforms, and Lin and Mielke (2008) reported the results from applying a mixture of Principal Component Analyzers on articulatory data to cluster place of articulation features.

2.5.3 The challenge of modelling phonological acquisition

In a review of models of phonological acquisition, Boersma et al. (2012) concludes that "there are no models yet that combine category creation to other emergent properties of

language processing, but that some partial answers have been given, so that we may well find a comprehensive model in the future." Indeed, many of the existing models offer ways of explaining various aspects of phonological competence, such as perceptual tuning, the acquisition of articulatory gestures, and how different layers of representation may interact. However, there are no models that can satisfactorily explain how auditory forms are transformed into abstract phonological representations in the lexicon.

2.6 Towards a model of phonological acquisition

This chapter offers a discussion of phonological representation that integrates linguistic theory and experimental findings from phonetics, phonology, and developmental psycholinguistics. Since the study of language acquisition has largely been an interdisciplinary effort, it is necessary to tie together various lines of research to better define the scope of the problem of phonological acquisition. The major themes of discussion in this chapter include the content and theories of phonological representation and the distinction between phonetic and phonological knowledge in first language acquisition.

With a comprehensive overview of the various aspects of phonological representation and acquisition, we can proceed to test specific hypotheses of how phonological acquisition proceeds. In this dissertation, I adopt the view that phonological features are emergent, and the goal of phonological acquisition is to arrive at a set of discrete lexical representations that best distinguish the lexical contrasts within the learner's vocabulary.

It is clear that the relationship between the continuous and variable speech signal and discrete phonological representations is by no means straightforward. In the next three chapters, I will address the following research questions:

1. Is lexical contrast a *sufficient* cue for the emergence of phonological categories?
2. Is there developmental evidence that the learner uses lexical cues in phonological acquisition?
3. How do we account for stability and sound change in an emergent phonological system?

Chapter 3

A Lexical Contrast Model of Phonological Acquisition

If we take phonological units as a set of symbols that can be used combinatorially in lexical representation, models of phonological acquisition should aim to satisfactorily explain how such symbolic units emerge. This chapter presents a model of phonological acquisition that accounts for the simultaneous learning of abstract phonological categories, their mapping onto the relevant acoustic features, and symbolic lexical representations using the acquired phonological units. This learning model introduces a mechanism of phonological category creation and refinement without the assumption of innately available phonological features. Central to this model is the idea that the need to represent lexical contrast is the driving force behind the creation and adjustment of phonological categories. The model, like the infant learner, begins with no phonological knowledge. As the model acquires words with distinct meanings, the need for abstract representation arises, and the model creates phonologically meaningful contrasts within the acoustic space to allow appropriate representations of the words in the learner's lexicon.

3.1 Lexical contrast and phonological acquisition

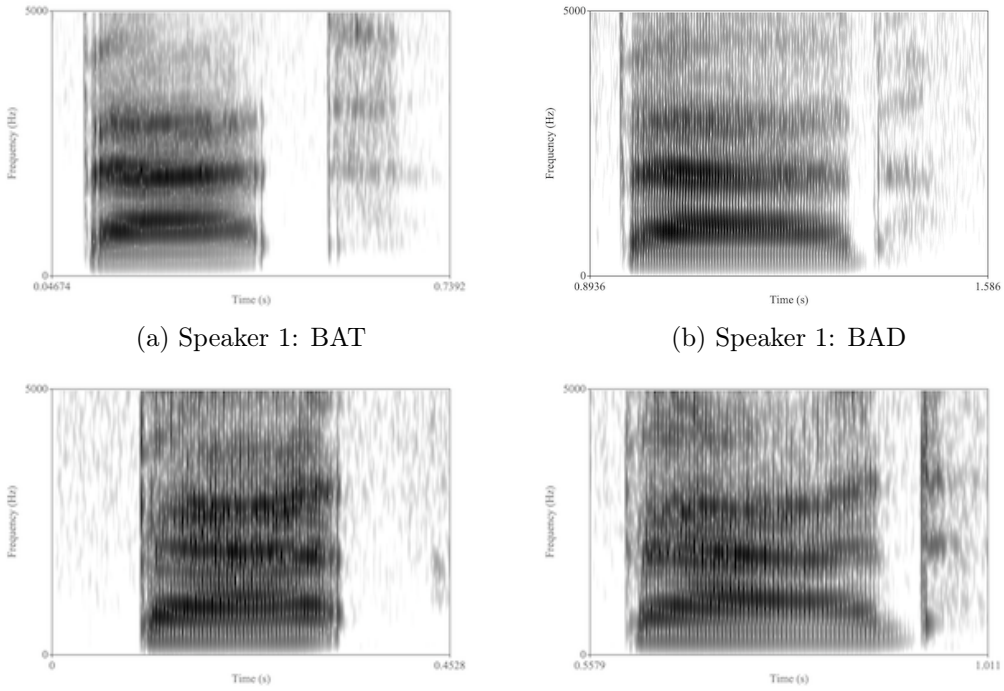
The notion of lexical contrast has a long history in phonology and was especially important in early approaches in phonology although it has received less attention in recent years (see Dresher, 2016, for a review). In phonological analysis, phonological distinctions are diagnosed via lexical contrast through the minimal pair test. More recently, researchers

in language acquisition have given word learning a more central role in the acquisition of phonological knowledge (Jusczyk, 1997; Werker and Curtin, 2005). This section reviews and discusses the importance of lexical contrast in phonological representation and offers motivation for a path of acquisition through the continuous restructuring of the phonological space to accommodate lexical distinctions.

3.1.1 Minimal pairs and lexical contrast

Phonological analysis operates on the symbolic level, which rests on the identification of abstract units of representation. Minimal pairs are a very efficient way of doing so. A minimal pair is two words that have distinct meanings and differ by only one unit. The unit is often assumed to be a segment. For English, "bin" and "pin" can be used to establish that /b/ and /p/ are distinct segments, i.e., phonemes. In commonly used feature theories, /b/ and /p/ are also minimal in the sense that they differ by only one phonological feature [voice]. Words such as "shin" and "bin" are a minimal pair and differ by one phoneme, but /ʃ/ and /b/ differ by more than one phonological or articulatory feature. While /ʃ/ is a voiceless alveolar fricative, /b/ is a voiced bilabial stop. As such, [ʃ] and [b] would also be more acoustically distinct than [b] and [p]. Additionally, for languages with suprasegmental features, minimal pairs can be found with words that share the same segments but differ in other aspects of articulation, such as pitch or phonation.

What role do minimal pairs play in phonological acquisition? Approaches that emphasize phonetic learning view minimal pairs as unnecessary (Maye and Gerken, 2000) and favor statistical learning. This approach often draws heavily from the perceptual discrimination results. However, as discussed extensively in Chapter 2, although perceptual discrimination provides compelling evidence for early phonetic development on the perceptual level, these results do not necessarily map directly to the development of abstract phonological categories. In addition to understanding the developmental trajectory of the discriminatory abilities themselves, it is equally important to carefully consider whether and how phonetic discrimination is used by the learner to parse linguistic input.



(a) Speaker 1: BAT

(b) Speaker 1: BAD

(c) Speaker 2: BAT

(d) Speaker 2: BAD

Figure 3.1: Spectrograms of the minimal pair "bat" vs. "bad" by two speakers.

The picture becomes more complicated when the details acoustic realizations are taken into consideration. Take the minimal pair "bat" and "bad" in English. When transcribed phonemically, they are respectively /bæt/ and /bæd/. Based on the phonemic analysis of the adult grammar, one might expect a minimal pair-based learner to identify the last segment as distinct phonemes. However, the actual acoustics of the two words suggests that this process is far more involved. Figure 3.1 illustrates the complications from the acoustic signal. Figure 3.1a and Figure 3.1b are the minimal pair produced by speaker 1. As can be seen, the acoustic distinctions between these two words are far from minimal. First, the vowel of "bad" is longer (each pair is plotted on the same time scale). The closure for the /t/ in "bat" is longer than the /d/ in "bad", and "bat" has a stronger release than "bad". There is a small amount of voicing for the /d/ in "bad". Since multiple acoustic cues differ between these two words, how does the learner figure out which ones are relevant? It would not be unreasonable to hypothesize that vowel length is the distinction between these two words, rather than the final consonant. Tokens from a second speaker further illustrate the

challenge of learning from the acoustic signal. In Figure 3.1c, the final /t/ is unreleased. Figure 3.1d has more prevoicing and a fairly strong burst and release. Similar to speaker 1, the vowel in "bad" is longer than the vowel in "bat". Clearly, a minimal phonological contrast does not correspond to a minimal phonetic contrast both within each speaker and across different speakers.

What, then, can the learner abstract away from knowing that the signal for "bat" and the signal for "bad" have different meanings and sounds? From two words that are acoustically different and referentially different, there is enough evidence that some contrast between them needs to be represented. This information is not sufficient to pinpoint the exact nature of this contrast, but learner can make an initial hypothesis about what to represent from the signal. Perhaps vowel length would be identified as the contrastive feature between "bat" and "bad", if the learner happens to perceive duration as the most salient difference between these two words. Then, as the learner acquires from words with /æ/ or encounter /t/ and /d/ in other contexts, the learner can use the additional lexical knowledge to evaluate the hypothesis that vowel length is the distinctive feature between "bat" and "bad". The important takeaway from these observations is that while the phonologist knows that "bat" and "bad" are a minimal pair, the learner does not. All the information the learner has is that these two words sound different and mean different things.

If a difference in signal and a difference in meaning are the only cues necessary for learning contrasts, the learner does not require phonological minimal pairs to start acquiring phonological contrasts. It is really the notion of lexical contrast that is important here. The words "fish" and "dog" differ by all three segments in adult English phonology. However, if these are the only two words a learner knows, the learner only needs two abstract symbols to represent them and can assign some acoustic salient cues to each symbol. In this initial state of phonology of the learner, "fish" and "dog" would actually be a minimal pair since they differ in sound and differ by one phonological unit of representation. Indeed, the phonological abstraction of what is contrastive is only *as detailed as the learner's lexicon needs it to be*. Minimal pairs in adult phonology may not correspond to minimal pairs

in a developing phonology because these phonologies can be very different. The minimal pairs in adult phonology are the end result of generalizing lexical contrasts over the acoustic space. Although the learner does not require minimal pairs to begin phonological acquisition, minimal pairs are nevertheless essential to the eventual refinement of phonological categories. Minimal pairs in the input grammar are words of high phonological signal, and they can help the learner to better pinpoint the relationship between abstract phonological units and their surface phonetic distinctions.

3.1.2 Phonological representation and lexical access

The phonological representations of words are accessed in word recognition. In mature adult phonology, homophones should have the same underlying phonological units, and experimental evidence suggests that this is in fact the case. Lexical decision tasks with homophones and non-word homophones show that words are phonologically encoded in the lexicon and that phonological processing occurs in the word recognition process. Some of this evidence comes from visual word recognition. Early work by Rubenstein et al. (1971) suggests that phonological processing does occur in lexical recognition. When subjects are presented with a homophonous non-word (e.g., brane), the reaction time is slower than phonotactically legal non-words without homophones. The longer latency for homophonous non-words is interpreted as longer search time as a result of phonemic matching. A separate experiment with all real words show that there is also a word frequency effect; low frequency homophones have higher latency and lower accuracy. Additionally, homophones facilitate the access of semantically related items (e.g., rows for flower, chare for table) (Van Orden, 1987; Lukatela and Turvey, 1991). Even though these experiments used orthography, the results indicate that orthography is parsed into some abstract phonemic representation, resulting in the observed effects from phonological homophones.

In the acoustic domain, word recognition is clearly not solely based on acoustics but rather combines acoustic and contextual cues. Because of the close association between phonology and phonetics, it would be easy to assume that phonology provides the mapping

between acoustics and abstract forms. This is partially correct. Phonology is a function that combines all levels of information (phonetic, phonological, morphological, syntactic, semantic, and pragmatic) to produce an abstract representation. When listening to prose, subjects sometimes fail to identify words with a phoneme mispronounced, especially in word initial positions (Cole, 1973; Cole et al., 1978). The retrieval of words is highly dependent on context. Syntactic and semantic context play a role in lexical parsing (Marslen-Wilson, 1975; Marslen-Wilson and Welsh, 1978), and listeners struggle to identify words when they are removed from their conversational context (Pollack and Pickett, 1963). On the segmental level, phoneme identification is also associated with contextual predictability of the words they occur in (Morton and Long, 1976).

3.1.3 Early lexical representation and underspecification

Research in lexical acquisition shows that word learning begins early (Borden et al., 1983; Tincoff and Jusczyk, 1999; Bergelson and Swingley, 2012), and that infants are aware of phonetic details in familiar words (e.g., Jusczyk and Aslin, 1995; Swingley, 2005, 2009; Mani and Plunkett, 2010). However, not all phonetic details may be encoded as phonologically relevant by the learner (Van der Feest and Fikkert, 2015). When the nuances of perceptual identification are investigated, it appears that certain aspects of words are remembered better than others. For example, the stressed portion of the word is better represented. For bisyllabic words, 11-month-old French infants failed to recognize familiar words when the medial consonant was modified, but still recognized the words when the initial consonant was changed in manner or voicing (Hallé and de Boysson-Bardies, 1996). The stress pattern in English is different, and early perception reflects this difference. At 11 months, English-learning infants did not recognize familiar words when the initial consonant was modified, but tolerated modifications to the medial consonant (Vihman et al., 2004).

Another line of research suggests that early representation is more holistic than segmental. In production especially, word forms appear be represented more holistically early on, and often only salient details are retained (Ferguson and Farwell, 1975; Walley, 1993). A

number of studies suggest that early lexical representation may be phonologically underspecified (Hallé and de Boysson-Bardies, 1996). Moreover, young children process phonetic similarity on the syllabic level rather than phonemic level, and they are better at identifying items that share multiple phonemes than a single phoneme (Treiman et al., 1981; Walley et al., 1986). Also, children are more influenced by coarticulatory cues. For example, they rely more on vowel formant transitions in identifying fricatives than adults (Nittrouer and Studdert-Kennedy, 1987; Nittrouer et al., 1989).

3.1.4 Word learning and referent resolution

How young children learn the meaning of words is an important research question. Much like acoustic data, the signal for word-referent mappings is extremely noisy. Even nouns referring to concrete objects can be difficult to identify since many interpretations can fit the scene in which they are uttered. However, even at a very early stage of word learning, infants are able to identify the intended referents to their acoustic forms (Bergelson and Swingley, 2012; Mani and Plunkett, 2010; Tincoff and Jusczyk, 2012). Different mechanisms have been proposed to account for the acquisition of word-referent mapping. Mutual exclusivity (i.e., no two words can have identical meaning) can help constrain the learning of new words (Markman and Wachtel, 1988; Markman et al., 2003). Cross-situational statistics, through which the learner keeps track of common signal and objects across multiple scenes, offers one account for the learning of word-referent mappings (Smith and Yu, 2008).

There is a lot of active research in this area, but it is beyond the scope of this dissertation to address how referents are identified. The model described in the next section incorporates a random element in the acquisition of words, but it does not propose a mechanism through which the correct identification of the referent is achieved.

3.2 A model of phonological emergence

This section introduces a concrete mechanism whereby the learner acquires discrete phonological representations from continuous, variable acoustic signal. Given a set of words in

a lexicon and their corresponding acoustic realizations, the model arrives at the relevant phonological features that best represent the contrasts in the lexicon. The two components of the model are the lexicon and its associated phonology. The lexicon stores each word's phonetic representation including exemplars, frequency, and its abstract representation according to the current state of the learner's phonology. The learner's phonological knowledge describes the relationship between acoustic cues and abstract phonological categories. For each phonologically contrastive dimension, the phonological knowledge enables the learner to transform the acoustic signal into abstract representations by paying attention to the cues that are informative for each contrast. At the end of learning, the model acquires 1) the appropriate number of phonological contrasts that are best suited to represent the lexicon, 2) which acoustic cues matter for each contrast, and 3) the abstract symbolic representation for each word in the lexicon.

This section describes the components and operations of the model and discusses the emergent properties of the model. To fully validate the model, the results from a computational experiment using acoustic data extracted from the Philadelphia Neighborhood Corpus is presented in the following section.

3.2.1 Lexical learning

Lexical learning begins early and forms the foundation of phonological learning (cf. Section 3.1.3). In this model, the lexicon module stores information about words that the learner has been exposed to. The learner keeps track of three pieces of information for each referent: its average (i.e., prototypical) acoustic signal, phonological representation, and frequency. The structure of the lexicon is illustrated in Figure 3.2.

The learner begins with no words in the lexicon. At each learning iteration, the learner is presented with the referent of a word and its acoustic signal. The model assumes that the learner is always able to correctly identify an acoustical signal with its referent, as in a perfect lab learning situation. The mapping between the signal and its referent is by no means a simple problem in language acquisition, but it is not a problem that this model

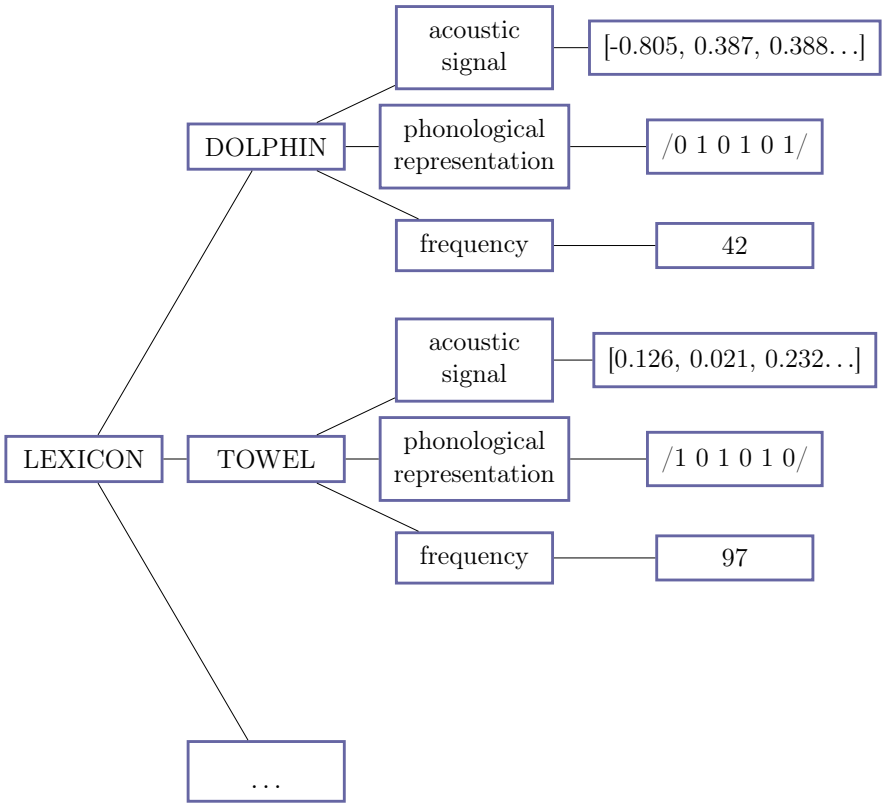


Figure 3.2: The structure of the lexicon.

aims to solve. As acoustic tokens for each referent are presented, the learner begins building up their knowledge of the phonetic forms that are associated with each referent. Since this model is primarily concerned with phonological acquisition, I make simplifying assumptions about the representation of a word's syntax, semantics, and pragmatics. The phonological learning part of this model only requires the learner to identify words as distinct in meaning along any of the dimensions of linguistic contrast.

The phonetic knowledge part of the lexicon reflect the learner's overall experience with phonetic forms of a word, and it includes any acoustic cue that the learner perceives from the input, both phonologically relevant cues and cues that do not contribute to any phonological contrast in the language. This phonetic knowledge is represented as the average of all the acoustic realizations corresponding to a referent, and it is updated each time an acoustic token for a referent is heard. As a result, after hearing a number of acoustic realizations

identifying a referent, the learner knows what a typical realization sounds like for this referent, and this process effectively creates an acoustic prototype for the phonetic realization of a word. After each iteration, the acoustic knowledge according to Equation 3.1, and frequency is updated according to Equation 3.2.

$$s = \frac{s \times f + s_i}{f + 1} \tag{3.1}$$

$$f = f + 1 \tag{3.2}$$

where:

f = word frequency; the number times a word has been heard
s = the existing prototypical (average) signal of a word
s_i = a specific acoustic token of the word

Before a word can make an impact on phonological learning, the learner needs enough familiarity with the word to be able to recognize it consistently. To simulate the increasing familiarity with a word with exposure, a simple frequency-based memory system is used to model the acquisition of words. The more frequently a word has been heard, the more likely that it is acquired by the learner and used in phonological learning. Before a word is acquired, the learner only updates their knowledge of the word on the phonetic level, and its phonological form is determined at the point of word acquisition. The acquisition of phonological contrasts and representations will be discussed in the following section.

The acquisition of a word is implemented as a probabilistic process with the likelihood increasing as the frequency of the word increases. After each token is heard, a random acquisition threshold t is generated from a uniform distribution between 0 and 1 (Equation 3.3). A random threshold is used to implement some noise in the learning process. The familiarity of a word is modeled as a logistic function (cf. Anderson et al., 1998) in Equation 3.4 (illustrated in Figure 3.3 for $k = 20$). If the familiarity r of the word is greater than the threshold t, the word is marked as acquired and pass onto the phonology module to be

assigned a phonological representation.

$$t = \text{unif}(0, 1) \tag{3.3}$$

$$r = \frac{1.0}{1.0 + e^{-(f-k)}} \tag{3.4}$$

where:

t = threshold at which a word is considered acquired

r = familiarity to the word

k = the word frequency at which $r = 0.5$

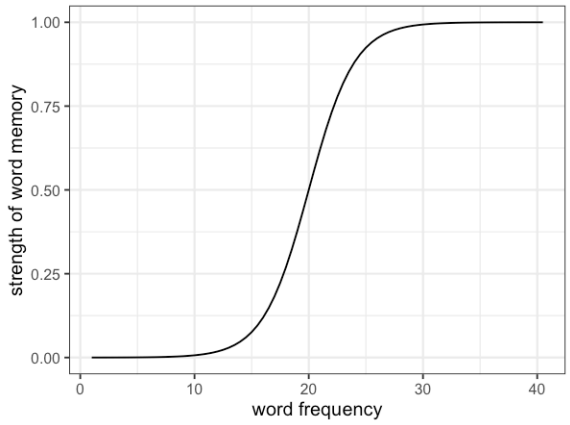
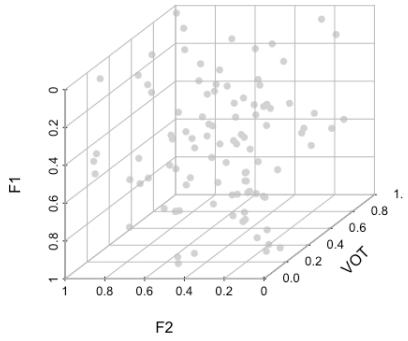
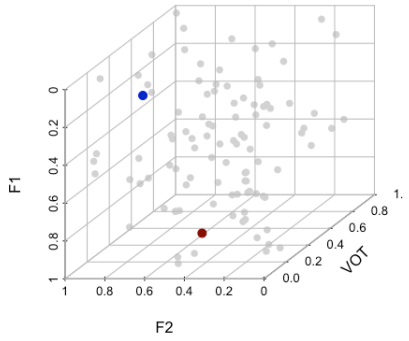


Figure 3.3: The probability of word familiarity as a function of word frequency.

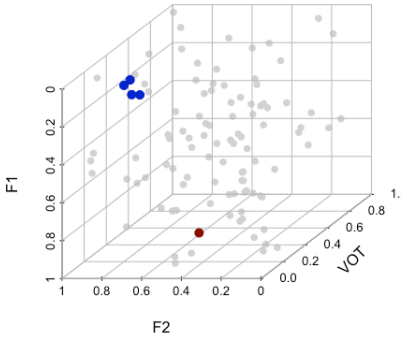
Figure 3.4 illustrates this process of word learning. These illustrations assume a toy language with only three acoustic dimensions (VOT, F1, F2) on the phonetic level and an unknown number of words. Figure 3.4a represents the stage prior to any lexical learning, and each grey dot represents some acoustic token of the words in this language. In Figure 3.4b, the learner begins paying attention to certain words, as represented by the BLUE and RED dots. Dots of the same color represent acoustic tokens that have the same referent. In Figure 3.4c, the learner is exposed to more tokens of BLUE. After some amount of exposure, the learner acquires BLUE, as represented by the big BLUE dot in Figure 3.4d). Further lexical



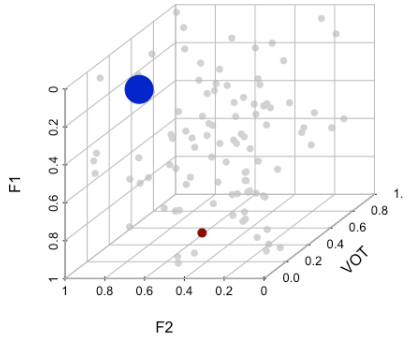
(a) The learner begins with no phonological contrast.



(b) The learner begins word learning.



(c) The learner hears many tokens of a BLUE.



(d) The learner acquires BLUE.

Figure 3.4: An illustration of lexical acquisition.

acquisition occurs the same way. After the learner hears tokens of the same word multiple times, the learner acquires this word and can use this word in phonological acquisition.

3.2.2 Phonological learning

Phonological learning occurs as the learner continuously makes hypotheses about how to transform the phonetic signal into abstract phonological categories that best represent the current lexical distinctions in the learner's lexicon. The learning is unsupervised and non-parametric; the learner does not know which phonological distinctions exist in the input

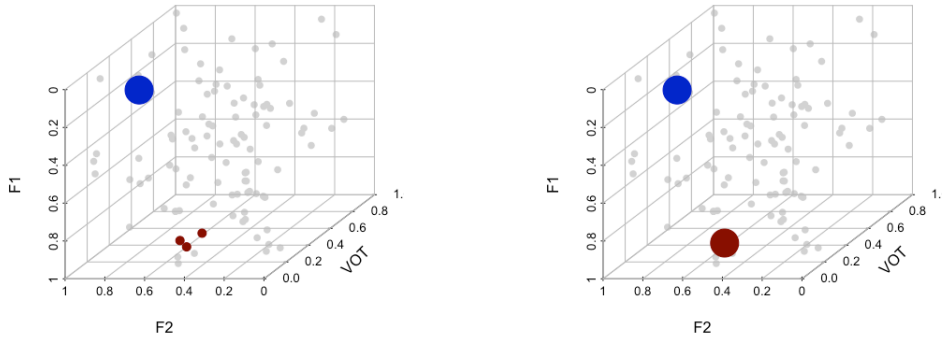
and is not given target representations. The learner's representations of words are updated dynamically as the learner acquires words and phonological contrasts.

The phonological module of the model consists of three processes: contrast creation, contrast adjustment, and contrast consolidation. In contrast creation, the learner adds a phonological contrast when the current number of contrasts is insufficient for representing the lexicon. After its initial creation, each contrast is updated as more words are learned and assigned to either side of the phonological boundary. Finally, should two contrasts become functionally the same after updates, they are consolidated into one contrast.

3.2.2.1 Contrast creation

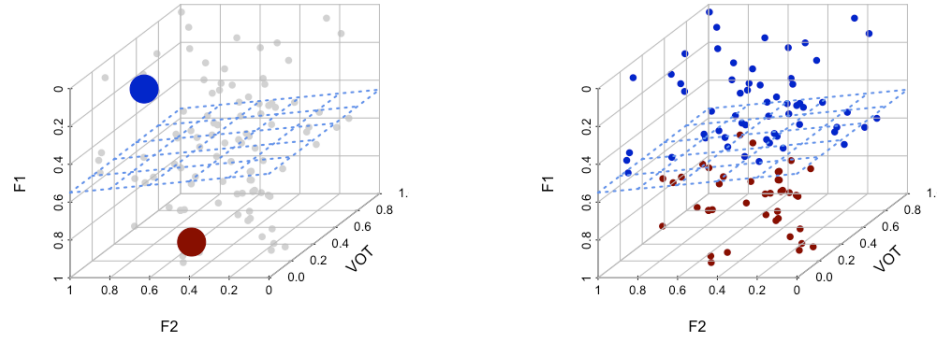
After a period of lexical learning, the learner will begin to recognize familiar words. When the learner acquires two words that are distinct in meaning, the learner needs to create the first phonological contrast that allows them to represent these two words distinctly. This is illustrated in Figure 3.5b, where the learner has acquired both BLUE and RED. To create the first contrast, the learner creates a division in the phonetic space that separates these two words based on the salience of the acoustic cues that distinguish these two words. The light blue plane in Figure 3.5c represents phonological CONTRAST #1, created after the learner has acquired BLUE and RED. Since these two words appear to be most distinct in F1, the plane cuts through the acoustic space mostly along the F1 dimension, with some tilt along the F2 dimension. The learner will be able to represent any subsequent acoustic tokens along this contrastive plane (Figure 3.5d). If the learner identifies another pair of words as distinct in meaning but current phonology represents them in the same way (BLUE and PURPLE in Figure 3.6a), the learner can create an additional contrast (the mostly vertical plane CONTRAST #2) to accommodate this need for distinct representation (Figure 3.6b). The number of phonological contrasts grows as the learner gains more vocabulary.

In the computational implementation, the learner's phonological knowledge is represented as a matrix W, where each column corresponds to an acquired phonological plane that divides the multidimensional acoustic space (Equation 3.5). At the beginning of learn-



(a) The learner begins learn a second word.

(b) The acquires a RED.



(c) The learner creates a phonological contrast in the acoustic space.

(d) The learner can use this acquired contrast to classify any token in this acoustic space.

Figure 3.5: An illustration of phonological contrast creation.

ing W is empty. Upon acquiring the first two words, the first phonological contrast is created. To create this contrast, the model compares the acoustic signals of the two words and determines the most acoustically salient cues between the two words. The relative salience of cues is calculated as the absolute value of the differences between each cue of the two words. Then, a phonological contrast is constructed as the plane equidistant from the most distinctive acoustic cues in the two words (Equation 3.6). Subsequent phonological contrasts are created in the same fashion, and phonological representations are assigned to each word using sigmoidal activation (Equation 3.7).

$$W = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \cdots & w_{m,n} \end{pmatrix} \tag{3.5}$$

$$\begin{aligned} W_{2:m,j} &= a_1 - \frac{a_1 + a_2}{2} \\ W_{1,j} &= -W_{2:m,j} \cdot \frac{a_1 + a_2}{2} \end{aligned} \tag{3.6}$$

$$p = \frac{1.0}{1.0 + e^{-Ws_i}} \tag{3.7}$$

where:

W = a matrix where each column is a phonological division in the acoustic space

$W_{1:m,j}$ = weights for the jth phonological contrast

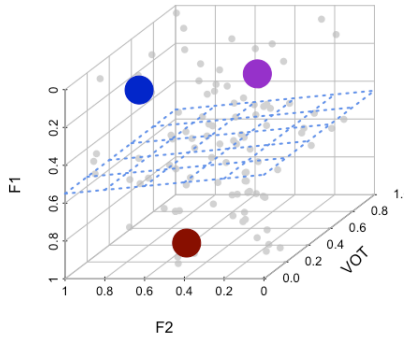
a_1, a_2 = the acoustically salient part of the signals of two distinct words

s_i = the acoustic signal from some word

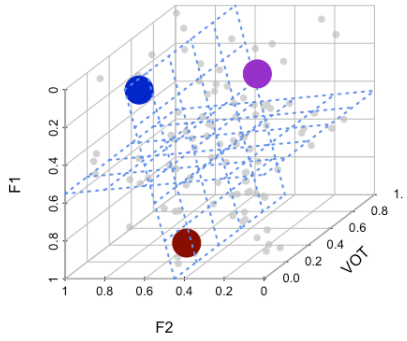
p = the phonological representation

3.2.2.2 Contrast update and adjustment

In addition to creating more phonological distinctions to represent the growing vocabulary, the phonological planes can also shift to to distinguish newly acquired word distinctions. This operation can be observed in Figure 3.7. In 3.7a, a new word, ORANGE has been acquired, and it falls in the same phonologically delineated space as PURPLE. In 3.7b, the

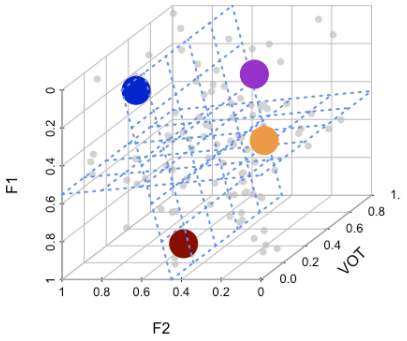


(a) The learner begins learn a third word PURPLE.

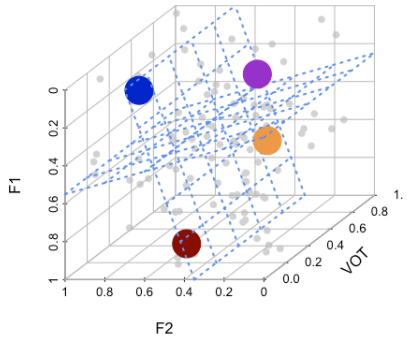


(b) The learner creates a second contrast.

Figure 3.6: The number of contrasts increases to accommodate the bigger vocabulary size.



(a) The learner acquires a new word ORANGE.



(b) The learner adjusts a phonological contrast to accommodate the lexicon.

Figure 3.7: The number of contrasts increases to accommodate the increased vocabulary size.

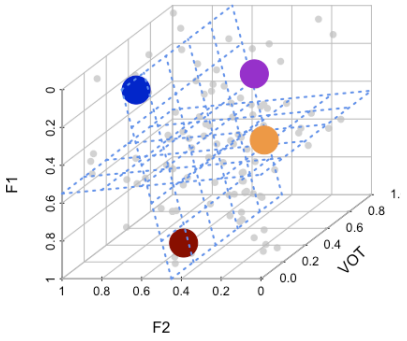
existing horizontal CONTRAST 1 tilts upward to phonologically separate PURPLE and ORANGE in the acoustic space.

As new tokens of existing words are heard and as new words are acquired and assigned phonological representations, all contrastive planes shift to best reflect the acoustic distinctions of the words assigned to either side of each boundary. For example, in 3.7b, there is also a slight shift in the vertical CONTRAST 2. The shift is the result of ad-

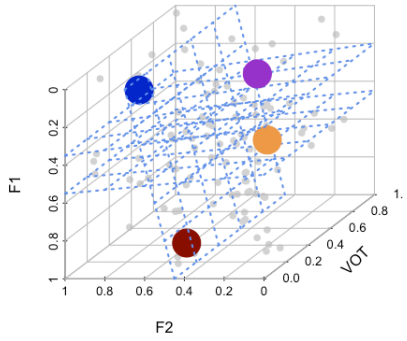
justing to the opposition of RED+BLUE vs. **PURPLE+ORANGE**, rather than just RED+BLUE vs. **PURPLE** (cf. Figure 3.6b). The plane is updated using Equation 3.6, where a_1 =mean(RED, BLUE) and a_2 = mean(PURPLE, ORANGE).

3.2.2.3 Contrast consolidation

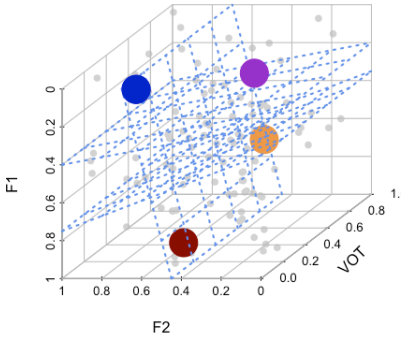
Because phonological contrasts are created based on prominent acoustic features of specific words, these contrasts can be word-specific initially. As more words are learned and contrasts become generalized across more lexical items, it is possible for two contrasts to become more and more phonologically similar. This scenario is depicted in Figure 3.8. Upon learning ORANGE 3.8a, rather than adjusting the boundary as in Figure 3.7, another possibility is that the learner creates an additional contrast as in Figure 3.8b. After learning more words (not represented in the plots to avoid visual clutter) and updating the boundaries, it is possible for two categories to become functionally equivalent. Illustrated in Figure 3.8c, both horizontal planes that create divisions mostly along F1 separate RED+ORANGE from BLUE+PURPLE. Because these two contrasts are functionally the same in this lexicon, they consolidate into one contrast (Figure 3.8d). In this case, consolidating the categories does not affect the system of contrast within the lexicon: BLUE remains distinct from RED, and PURPLE remains distinct from ORANGE. The developmental interpretation for this consolidation of categories is that learners tend to learn word-specific contrasts initially. The learner might acquire a contrast /b/ vs. /d/ from "ball" and "doll", then acquire a similar contrast /b'/ vs. /d'/ from "boo" and "do" because the phonetic realizations of /b/ and /d/ might be different as the result of coarticulation with the following vowel. As the learner acquires more words and adjust the phonological boundaries, word-specific phonetics will be attenuated, and /b/ vs /d/ and /b'/ vs /d'/ will become more similar and eventually consolidated as the same categories.



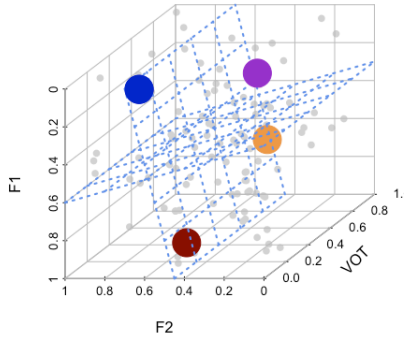
(a) The learner acquires a new word ORANGE.



(b) The learner creates another phonological contrast.



(c) The two contrasts become functionally the same.

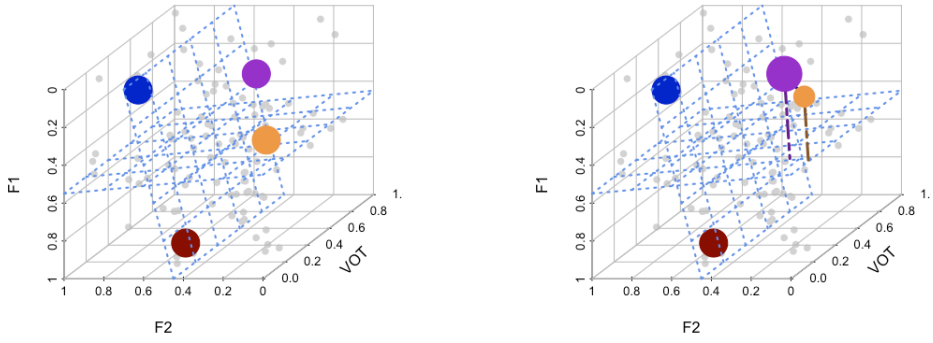


(d) The two contrasts consolidate.

Figure 3.8: An illustration of phonological contrast consolidation.

3.2.2.4 Contrast determination

The above presents two mechanisms that two words can be represented as distinct. The model can create a new phonological contrast or adjust an existing contrast to accommodate the increasing lexical distinctions that need to be represented. However, homophones exist in language, and mergers as a sound change are very common. A model of phonological acquisition should be able to account for the existence of true homophones. How does the model choose between 1) creating a new contrast, 2) adjusting an existing contrast, and 3)



(a) The learner acquires a new word ORANGE.

(b) ORANGE is less frequent than PURPLE.

(c) ORANGE is acoustically similar to PURPLE.

(d) The learner acquires a new word ORANGE.

Figure 3.9: An illustration of phonological contrast generalization and merger.

representing two words as homophones?

How does the learner conclude which items in their lexicon are better represented homophones? The choice depends on the acoustic distance between the two words in question, the existing phonological contrasts, and the relative frequencies of the two words. The motivation for this decision comes from psycholinguistic findings about lexical access. When two words have the same phonological form, the more frequent of a homophonic pair is accessed first regardless of syntactic and semantic context (Boland and Blodgett, 2001; Caramazza et al., 2001; Bonin and Fayol, 2002). The processing cost of representing two words as

homophonous can thus be quantified as the relative frequencies of the two words.

Returning to the learning stage where ORANGE has just been acquired (Figure 3.9a). There are possible scenarios it might be advantageous for the learner to represent PURPLE and ORANGE as homophones rather than representing them distinctly. For example, if ORANGE is less frequent than PURPLE and the two words are acoustically close (Figure 3.9b), a learner that assigns the same representation to PURPLE and ORANGE would still correctly identify PURPLE as the intended referent most of the time. If the intended referent is ORANGE, the learner would access PURPLE first and need additional processing to access the less frequent form ORANGE. This delay in processing can be quantified using the frequencies of the two words. With homophonous representations, the delay in processing from representing the two words as homophones can be quantified as follows:

$$C_{homophone} = \frac{\text{freq(ORANGE)}}{\text{freq(ORANGE)+freq(PURPLE)}} \tag{3.8}$$

On the other hand, if ORANGE and PURPLE are more acoustically distinct (Figure 3.9d), it might make sense to represent them distinctly even if PURPLE is far more frequent. Two factors need to be considered in making this determination. First, are PURPLE and ORANGE sufficiently acoustically distinct to warrant the creation of a new contrast? Second, is ORANGE frequent enough to warrant a distinct lexical representation? The first factor can be quantified using a measure of acoustic confusability between the two words:

$$\text{confusability} = \frac{\text{d(PURPLE, boundary)}}{\text{d(PURPLE, boundary)+d(PURPLE, ORANGE)}}$$

where:

$$d(a_1, a_2) = \sqrt{\sum_{i=1}^{m}(a_{1i} - a_{2i})^2}$$

The closer ORANGE is to PURPLE, the more acoustically similar they are. If they are too acoustically similar, creating a contrast between them will likely result in confusion in perception. This confusability measure is calculated based purely on acoustics, and it is still necessary to take into account the relative frequencies of the two words. If both words have the same frequency, they would be confused with each other by this measure. However, since the two words are not equally frequent, a weighted confusability measure can be used to quantify the processing cost of having contrastive representations:

$$C_{contrastive} = \frac{\text{freq(PURPLE)}}{\text{freq(ORANGE)+freq(PURPLE)}} \times \text{confusability}$$

If the processing cost of homophonic representation is greater than contrastive representation ($C_{\text{homophone}} > C_{\text{contrastive}}$), the learner either adjusts existing contrasts or create a new contrast to be able to represent these two words distinctly. Otherwise, homophonic representations are tolerated.

Lastly, one more scenario is illustrated in 3.9d, where shifting the existing phonological plane to distinguish PURPLE and ORANGE would make ORANGE homophonic with GREEN. Therefore, creating a new contrast would be the only option here if the learner determines that PURPLE and ORANGE need to be represented distinctly.

3.2.3 Emergent representations and properties of the model

This learning mechanism outlined in this section has several emergent properties, which are discussed below.

Phonological features. Some prominent treatment of phonology assume innate, universal phonological features (Jakobson, 1968; Chomsky and Halle, 1968; Reiss, 2018). This learning model illustrates a concrete path by which phonological features can be acquired using only acoustic and lexical cues. There is no need to assume innate phonological features. Some abstract category formation mechanism would be sufficient, either domain-general or

guided by UG.

Acoustic cues for phonological features. While learning phonological categories, the model simultaneously learns the mapping between these categories and the relevant acoustic cues. By comparing the acoustics of lexical items, the model identifies which acoustic cues are meaningful to a phonological contrast and their relative contribution to the identification of the phonological contrast.

Discrete lexical representations. Discrete lexical representation are assigned to each word as soon as it is acquired. The creation of phonological boundaries enables to learner to transform the acoustics of each word into phonological distinctions.

Increasing specificity of lexical representation. The learning mechanism naturally address early underspecification that has been reported by many studies (Hallé and de Boysson-Bardies, 1996; Vihman et al., 2004; Fikkert and Levelt, 2008). The lexical representations themselves become more specified when more words are learned, and the differences between infant and child language can be largely explained in terms of the size of the vocabulary. With few words, the apparent underspecification can come from two sources. First, the learner does not need as many symbols to represent fewer words, leading to the generalization of more phonetic information over fewer symbols. Second, with a smaller vocabulary, the learner may be inaccurate in determining which specific acoustic cues matter for a phonological contrast or fail to compensate for coarticulatory effects.

Minimal pairs. Because phonological representations are built on lexical contrast, minimal pairs arise naturally as the result of the learning process.

Feature economy. Feature economy refers to the idea that languages tend to maximize the use of contrastive dimensions (Clements, 2003). Because phonological contrasts are only created as needed from lexical and acoustic cues, the resulting system is naturally economical. As more words are acquired, more dimensions of contrasts are learned, but the

growth of contrasts is much slower than the growth in vocabulary.

3.2.4 Advantages of the model

A general approach to phonological acquisition. This is a general and integrated model for phonological learning and aims to learn any phonological contrasts. While many computational models focus on specific contrasts and only use cues for the contrast in question, this model makes use of the acoustic information over an entire word to learn cue weighting, and abstract lexical representations simultaneously.

Minimal theoretical assumptions. The applicability of this model is not dependent on existing phonological frameworks. The abstract representations in learned through the model can be used for further phonological analysis.

Minimal memory requirement. Because the learning is online, this model does not require calculations over a large number of input items. This model only requires the learner to remember the general acoustic shape of each word, their phonological representations, and the cue weights for each learned phonological contrast.

Non-parametric learning. This model is completely unsupervised and nonparametric. The learner does not know what contrasts exist and which cues matter for particular contrasts, both of which are discovered in the learning process. Also, the learning result is consistent and not dependent on the initialization of parameters. Third, this model can makes use of dynamic and overlapping acoustic information in word learning.

3.3 Experiment

The learning mechanism described in Section 3.2 is implemented computationally to test its validity. Acoustic measurements are extracted from the Philadelphia Neighborhood Corpus as input to the model, and the learning outcomes for phonological contrasts, acoustic cue weights, and lexical representations are presented.

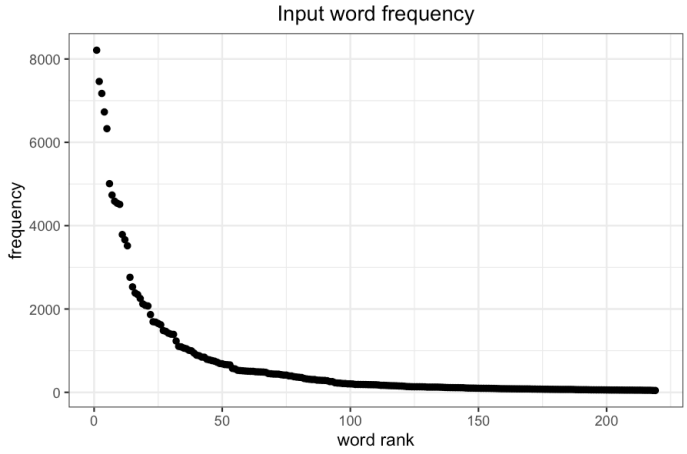


Figure 3.10: Input word frequencies.

3.3.1 Input preparation

Most of the previous work in the computational modeling of phonological/phonetic acquisition use simulated data as input (e.g., Vallabha et al., 2007; Feldman et al., 2013a). In order to better represent the noisy data that the child learner is faced with, this study uses real acoustic measurements from the Philadelphia Neighborhood Corpus (Labov and Rosenfelder, 2011). The input is limited to monosyllabic words with the syllable structures V, CV, VC, and CVC. Words containing nasal segments were excluded because of difficulty with automatically tracking measures of nasality across a large number of speakers. Words with frequencies 20 or fewer in the entire corpus are omitted.

3.3.1.1 Measurement extraction

A Praat script was written to automatically extract measurements from the corpus. For each segment, measurements were taken at 25%, 50%, and 75% of the duration of the segment. For all consonants, duration, center of gravity, jitter, shimmer, HNR (harmonics-to-noise-ratio), and autocorrelation were extracted. For sonorant consonants, f0, F1, F2, F3, B1, B2, and B3 were also extracted. Most vowel measurements, including F1, F2, F3, B1, B2, and B3 are available with the PNC. An additional measurement f0 is extracted for vowels.

3.3.1.2 Measurement normalization

Because the measurements were extracted automatically, normalization was carried out to replace potential tracking errors. The formant values were transformed onto the bark scale, and the f0 values were transformed onto semitones for each speaker. Measurements below 10% and above the 90% percentiles on the group level were changed to the group mean, and all the measurements were z-scored.

3.3.1.3 Descriptive statistics of the input

There are measurements from a total of 383 subjects from the PNC. Overall, there are 219 word types. Out of the word types, there are 162 CVC words, 30 CV words, 24 VC words, and 3 V words. There are 153,438 total word tokens, and 62909 CVC, 59934 CV, 28166 VC, and 2429 V word tokens. There are 16 onset phonemes (including null onset), 11 nucleus vowels, and 14 coda phonemes represented in the input data (including null coda). In total, 42 phonological oppositions are present among the phonemes in each position (Table 3.1).

Onset	anterior, approximant, back, consonantal, continuant, coronal, delayed release, distributed, dorsal, front, labial, labiodental, lateral, round, sonorant, strident, voice
Nucleus	back, diphthong, front, front.diphthong, high, labial, long, low, round, stress, tense
Coda	anterior, approximant, consonantal, continuant, coronal, delayed release, distributed, dorsal, labial, labiodental, lateral, sonorant, strident, voice

Table 3.1: Actual phonological contrasts in the input words for each position.

3.3.1.4 Representation of the input

Each segment of a word is represented as a 14-element vector with the measurements in the follow order: phoneme duration, f0, F1, F2, F3, B1, B2, B3, center of gravity, voicing, jitter, shimmer, autocorrelation, HNR. If a segment is null (for instance, for VC words the

onset is null), a vector containing 14 0's is used. Each instance of a word is represented as a 42-element vector (14 cues × 3 segments).

3.3.1.5 Learning trials

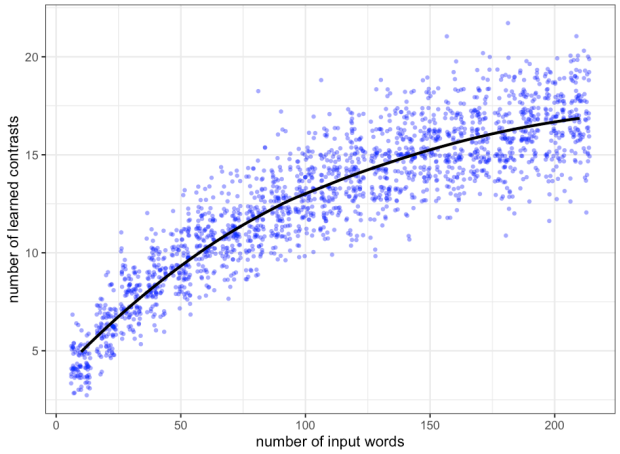
In total, there were 2100 trials with 21 input vocabulary sizes in increments of 10 (10 words, 20 words, 30 words, etc.) and 100 learning trials for each input vocabulary size. That is, for 100 trials, the model randomly picks 10 out of the 219 word types and uses the acoustic tokens of these 10 word types as input for learning. For each trial, 10 different words are randomly sampled. After 100 trials with 10 input words are terminated, 100 trials with 20 random input words are run, and so on. Learning is terminated after the number of phonological contrasts has stayed stable for 20,000 iterations. To evaluate the learning process and outcome of the learning model, the learned phonological weights, lexical representations, and word frequencies are logged every time there is a phonological change (i.e., addition or consolidation of phonological contrasts) and also every 1000 iterations.

3.3.2 Results

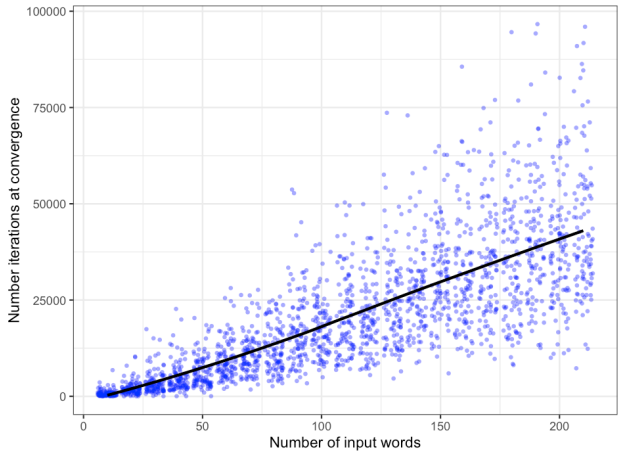
Overall, the model learns reasonable numbers of categories for the number of input words, and phonological contrasts converged for all trials. Case studies of specific learning trials show that the learned representations and the acoustic cues approximate phonological features commonly used for phonological analysis.

3.3.2.1 The effect of input vocabulary size

Across 2100 learning trials with varying input sizes, the model learned more contrasts for larger numbers of words. Figure 3.11a plots the number of categories the model learned for all learning trials, and a numerical summary of the results is presented in Table 3.2. The effect of vocabulary size on the number of contrasts learned is apparent. As the number of input words increased, the model learned more phonological categories to represent the words that have been acquired in the lexicon. The growth of phonological contrasts appears to flatten



(a) Learned number of contrasts for all the trials by the number of input words.



(b) Number of iterations needed for phonological convergence by the number of input words.

Figure 3.11: Learning outcome as the number of input words increases.

out with more number of words. This behavior of the model is expected, since the theoretical minimum number of binary contrasts required to represent N words is $\log_2(N)$. For 210 words, the minimum number of contrasts needed is eight ($\log_2(210) = 7.71$). The model learns on average twice the number of the theoretical minimum for 210 words. This could be partially the result the actual number of contrasts that exists in the input words. When all input words are considered, there are 42 distinct features using a feature system proposed for phonological analysis (Table 3.1). Compared to the 42 actual distinctive features that can be identified from these words, 16-17 learned features is reasonable for the given vocabulary.

# input words	contrasts	sd
10	4.09	1.13
20	5.93	1.43
30	7.54	1.40
40	8.50	1.61
50	9.47	1.54
60	10.65	1.31
70	11.03	1.34
80	11.63	1.63
90	12.39	1.48
100	13.03	1.48
110	13.46	1.59
120	14.08	1.64
130	14.25	1.73
140	14.95	1.60
150	15.41	1.46
160	15.46	1.77
170	15.99	1.51
180	15.74	1.92
190	16.64	1.48
200	16.56	1.70
210	17.06	1.77

Table 3.2: Average number of phonological contrasts learned over 100 learning trials for increasing numbers of input words.

Figure 3.11b displays the number of trials needed before the model converges on a set of phonological contrasts. As defined in Section 3.3.1.5, phonological convergence is achieved when there have been no changes to phonological contrasts for 20,000 iterations. The average number of iterations needed for convergence increases as the number of input word increases, but the variance also becomes greater as the number of input word increases. With more words, the model needs to account for a wider range of phonetic variation. Because word learning is probabilistic, in some cases, the model might acquire more generalizable contrasts earlier, resulting in the lower number of iterations needed for convergence. It is also possible that the model will need to re-tune the phonological contrasts many more times before achieving a stable state, thus resulting in a greater number of trials needed before convergence.

3.3.2.2 An example of learned representations

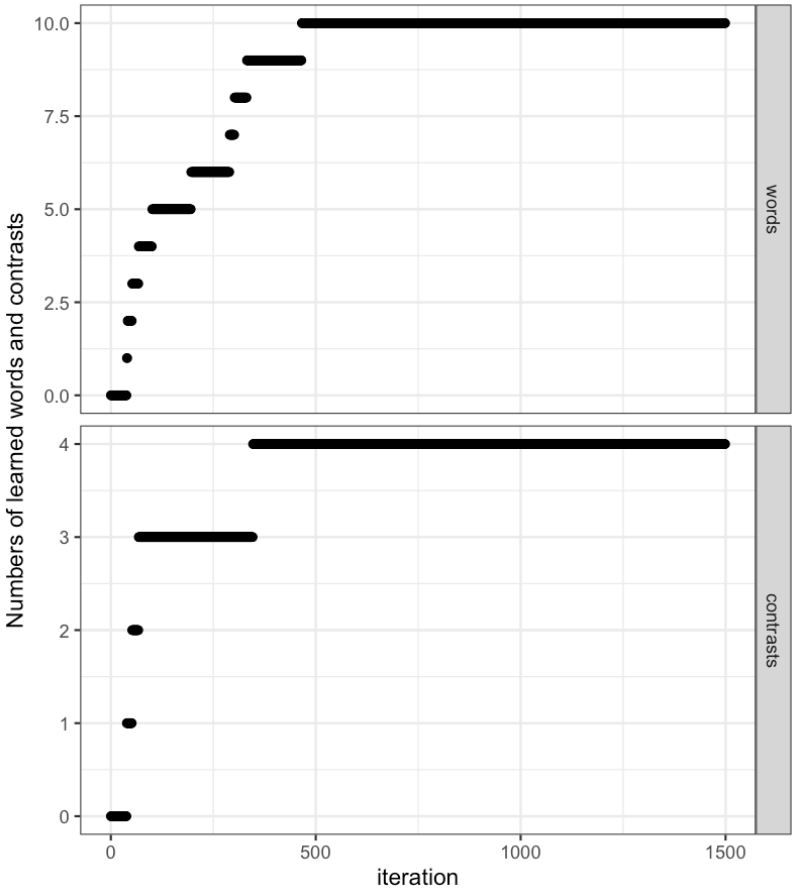


Figure 3.12: Word and contrast learning trajectories for a 10-word trial.

Each learning trial produces three results: the number of contrasts, cue weights for each contrast, and lexical representations based on these learned contrasts. This section presents a typical learned outcome from a learning trial with 10 words. The small number of words makes the results more easily interpretable. This particular instance of the learning outcome produced 4 phonological contrasts. The acquisition trajectory on the word level and the phonological level is illustrated in Figure 3.12. In this particular trial, a stable phonological state is reached on iteration 347, before all ten words have been acquired on iteration 466. The rest of this section will present the learning outcome of this trial by referencing the learned contrasts, learned representations, and comparisons to actual

contrasts from a phonological analysis of these words.

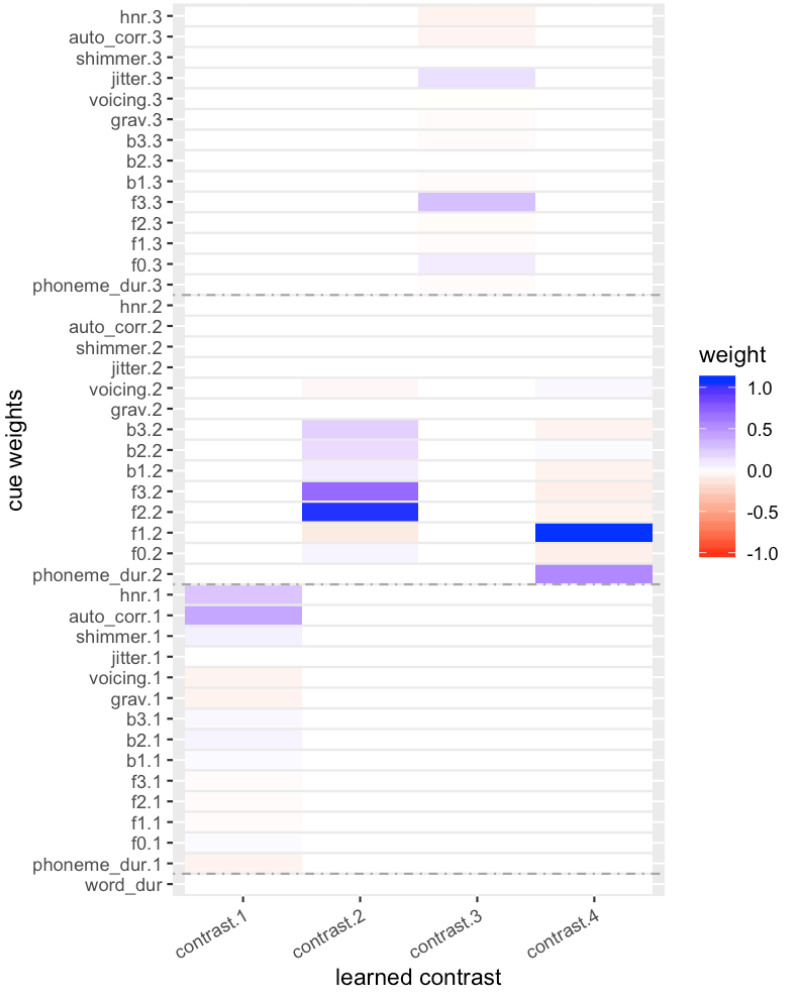


Figure 3.13: Learned weights for each of the four contrasts for a 10-word learning trial.

Figure 3.13 illustrates the learned cue weights in the form of a heatmap. Each column in the plot corresponds to one phonological contrast. Darker colors (either more blue or more red) indicate that the acoustic cue is more important for the contrast. Table 3.3 presents the learned lexical representations according to the four phonological contrasts. The use of "0" and "1" are purely symbolic and they merely indicate distinction along a phonological dimension. All the words that have the representation "0" fall on one side of the phonological division in acoustic space, while all the words with the representation "1" fall on the other. Which words are assigned "0" and which are assigned "1" is arbitrary. Figure

3.14 shows how the learned representations correspond to phonological features typically used in phonological analysis.

referent	contrast 1	contrast 2	contrast 3	contrast 4
FAR	0	0	0	1
ARE	1	0	0	1
OR	1	0	0	0
DEAL	1	1	0	0
WE'VE	1	1	1	0
FEEL	0	1	1	0
TOOK	0	0	1	0
COP	0	0	1	1
PAID	0	1	1	1
CAT	0	1	0	1

Table 3.3: Learned lexical representations with 10 words in the input.

According to Figure 3.13, the first contrast learned in this trial is an onset contrast, and the relevant cues are autocorrelation and HNR. These acoustic cues are typically associated with the manner or voicing of consonants. In the learned representations (Table 3.3), words with voiceless onsets (FAR, FEEL, TOOK, COP, PAID, CAT) are separated from words with voiced onsets or no onsets (ARE, OR, DEAL, WE'VE). Indeed, when comparing the learned contrast with phonologically analyzed contrasts, this dimension correlates highly with voicing and manner features (Figure 3.14). Moving on to the second learned contrast, the heavily weighted phonetic cues are in the nucleus, and the most important cue is F2 (Figure 3.13), which usually indicates differences in the frontness or backness of the vowel. Indeed, in the learned representations, this contrast marks the distinction between back vowels (FAR, ARE, OR, TOOK, COP), and front vowels (DEAL, WE'VE, FEEL, CAT), and the learned representations correspond to [front] and [back] features in traditional phonological analysis (Figure 3.14).

As for contrast 3, the phonetic weighting indicates that this is a coda contrast based on F3 differences (Figure 3.13). Unlike contrast 1 and contrast 2, this contrast does not correspond neatly to any phonologically analyzed contrasts. For the most part, words with sonorant codas (FAR, ARE, OR, DEAL) are separated from stop and fricative co-

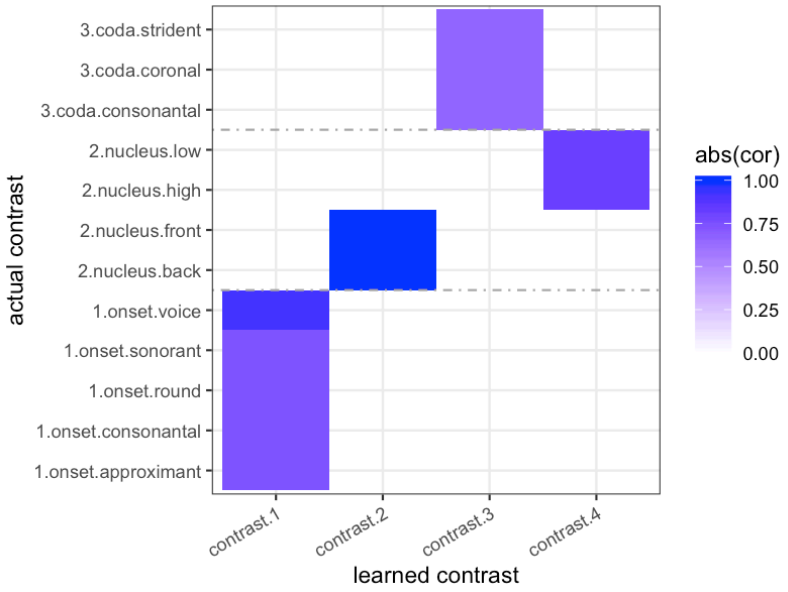


Figure 3.14: Correlation of learned representations to to actual phonological features for a 10-word trial.

das (WE'VE, TOOK, COP, PAID). However, CAT and FEEL do not fit this pattern. The learned representation for CAT is /0 1 0 1/. If CAT is represented as /0 1 1 1/, it would be homophonous with PAID. Perhaps this is the reason that the model adjusts this contrast to accommodate CAT vs. PAID in the existing phonological space rather than creating a new contrast. The assignment of FEEL to /1/ for contrast 3 is anomalous and may be the result of the specific acoustic measurements of FEEL. A more general voicing contrast may be acquired with more words. Finally, the last learned contrast distinguishes vowel height, with F1 as the most prominent acoustic feature. Table 3.3 shows that words with low vowels (FAR, ARE, COP, PAID, CAT) are separated from words with high vowels (OR, DEAL, WE'VE, FEEL, TOOK), and this is confirmed by the high correlations to manner features in Figure 3.14. There are minimal pairs in the learned phonological representations, but these minimal pairs are defined within the phonological contrasts learned from these 10 input words. FAR /0 0 0 1/ and ARE /1 0 0 1/ differ by the onset Contrast 1. This corresponds to the actual phonological contrast that FAR has an onset /f/ and ARE has null onset. The rest of these two words have the same representations. Similarly, ARE /1 0 0 1/ and OR /1 0 0 0/ form a minimal pair and differ only in Contrast 4, a vowel height contrast.

However, within this phonology, TOOK /0 0 1 0/ and COP /0 0 1 1/ are also a minimal pair even though in actual English phonology they differ by all three segments. These two words are fairly acoustically similar: They both have a voiceless stop in the onset and a voiceless stop in the coda. With a small vocabulary of 10 words, representing TOOK and COP as a minimal pair is entirely reasonable. The difference in the vowel – the distinctive part between these two words – is enough for the learner to identify the contrast between these two words within this small lexicon. The learner is being efficient (or economical) in this kind of use of their phonological space. With a larger vocabulary, the learner will need to create more fine-grained contrasts between the different stops, but this is not necessary given the acoustics and the lexical contrasts in the input of this trial.

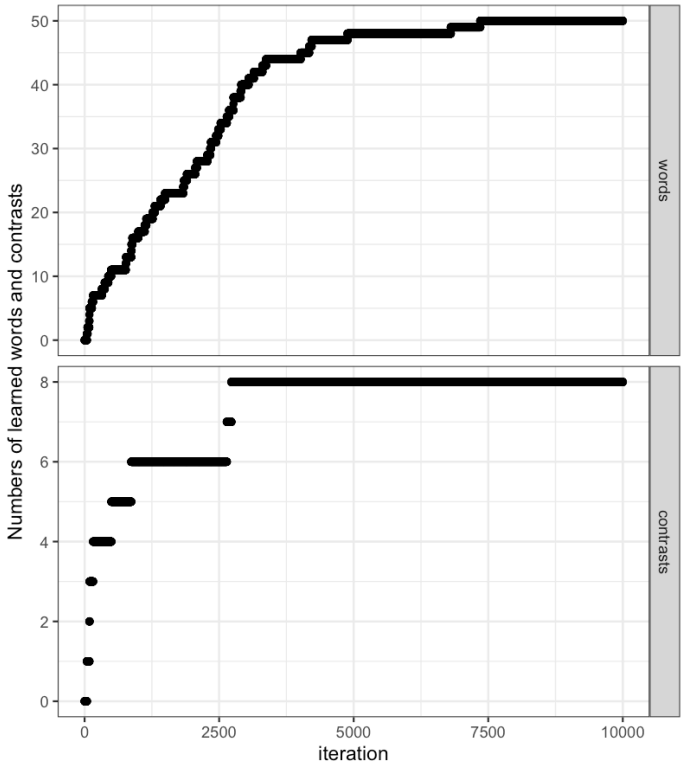


Figure 3.15: Word and contrast learning trajectories for the 50 word trial.

3.3.2.3 Learning outcome with 50 words

The model is successful at discovering meaningful contrasts for the 10-word trial presented above. Does this result generalize to the learning of more words? In this section, the results from a 50-word learning trial are presented. The learning trajectory for words and phonological contrast is shown in Figure 3.15. For this case, the number of phonological contrast stabilizes at iteration 2728, when 36 words have been acquired. These learned representations are sufficient to accommodate the words that have not yet been learned. All 50 words are acquired at iteration 7352.

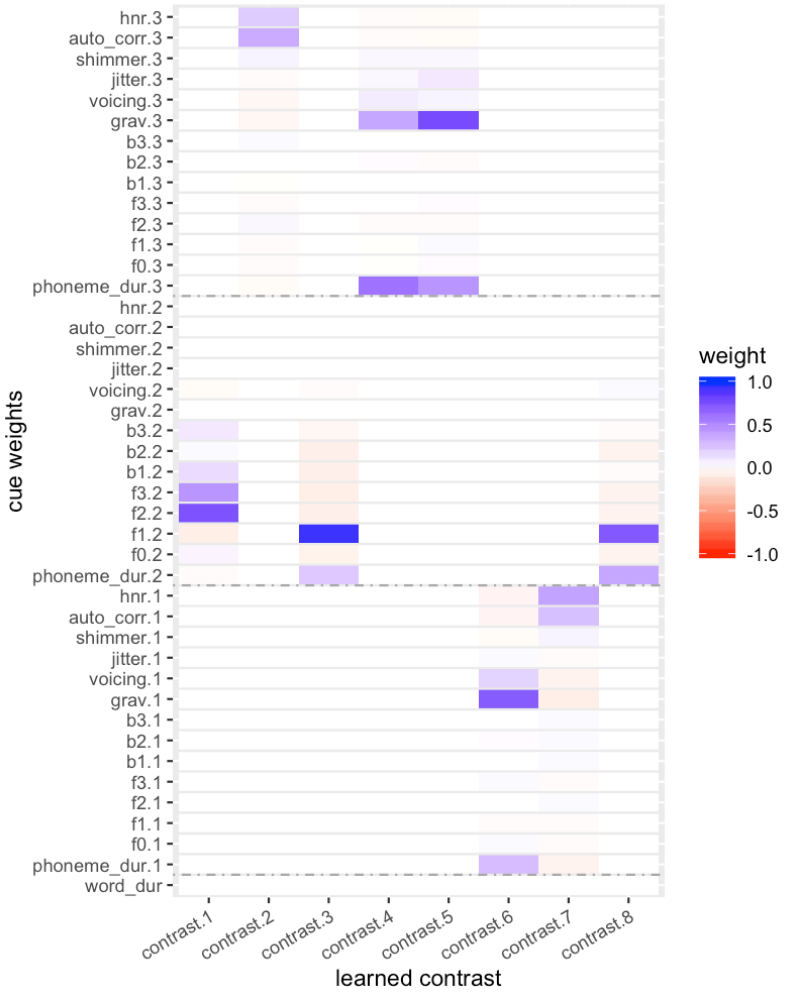


Figure 3.16: Learned contrasts for 50 words.

Figure 3.16 shows the learned cue weights. A total of 8 contrasts were learned, among

which there are two onset contrasts (#6 and #7), three vowel contrasts (#1, #3, #8), and three coda contrasts (#2, #4, #5). Since listing all the learned representations is not as easily interpretable as the 10-word trial, I will discuss the learned representation results on the segmental level.

	Contrast 6		Contrast 7	
	Fricatives + t		Voice	
	0	1	0	1
∅	100.00	0.00	0.00	100.00
b	100.00	0.00	0.00	100.00
p	100.00	0.00	100.00	0.00
d	100.00	0.00	50.00	50.00
t	0.00	100.00	100.00	0.00
g	100.00	0.00	25.00	75.00
k	50.00	50.00	100.00	0.00
f	33.33	66.67	100.00	0.00
s	0.00	100.00	100.00	0.00
ʃ	33.33	66.67	100.00	0.00
l	100.00	0.00	50.00	50.00
ɹ	100.00	0.00	0.00	100.00
w	100.00	0.00	0.00	100.00
j	100.00	0.00	0.00	100.00

Table 3.4: Percentages of each onset phoneme assigned to each side of a learned phonological contrast.

The learned onset distinctions are compared to actual phonemic representations in Table 3.4. For each learned phonological contrast, this table presents the percentages of the learned representations for each phoneme. For instance, /p/ is assigned /0/ for the learned Contrast 6 in 100% of the word types it occurs in, while /t/ is assigned /1/ for Contrast 6 in 100% of the word types it occurs in. According to Figure 3.16, Contrast 6 separates onset phonemes by the acoustic cue center of gravity. Comparing this to the assigned representations, it appears that Contrast 6 separates fricatives from the rest of the phonemes. The phoneme /t/ is grouped with the fricatives, possibly because its coronal place of articulation results in similar average frequencies as /s/ and /ʃ/. Contrast 7 is more straightforward; it creates a boundary between voiced and voiceless onsets by dividing the acoustic space mostly along

HNR and autocorrelation, both are measures of periodicity in the signal.

	Contrast 1 Front/Back		Contrast 3 High/Low		Contrast 8 High/Low	
	0	1	0	1	0	1
i	0.00	100.00	100.00	0.00	100.00	0.00
ɪ	100.00	0.00	100.00	0.00	100.00	0.00
e	0.00	100.00	0.00	100.00	40.00	60.00
ɛ	16.67	83.33	0.00	100.00	83.33	16.67
æ	66.67	33.33	0.00	100.00	0.00	100.00
ɑ	100.00	0.00	0.00	100.00	20.00	80.00
ʌ	100.00	0.00	0.00	100.00	50.00	50.00
ɔ	100.00	0.00	0.00	100.00	0.00	100.00
o	100.00	0.00	0.00	100.00	100.00	0.00
ʊ	100.00	0.00	100.00	0.00	100.00	0.00
u	85.71	14.29	100.00	0.00	100.00	0.00

Table 3.5: Percentages of each vowel phoneme assigned to each side of a learned phonological contrast.

The learned representations for each vowel is presented in Table 3.5. Contrast 1 separates the phonemes mostly along F2, which corresponds to the frontness or backness of the vowel. The acoustic boundary separates /i e ɛ/ from the rest of the vowels. This learned boundary appears to be very "front": The vowel /ɪ/ and /æ/ are typically described as front in phonological analysis, but they are grouped with the back vowels in this learned contrast. Contrast 3 clearly distinguishes high vowels from non-high vowels. Contrast 8 is a second high-low contrast, but the boundary appears to be "lower" than Contrast 3. Contrast 8 separates the mid vowels /o/ and /ɛ/ from the low vowels, but /e/ is ambiguously represented by this contrast.

There are three contrasts learned for the coda (Table 3.6). Contrast 2 corresponds to voicing and separates the voiceless codas /p t k s ʃ θ/ from the voiced ones /d v z ɹ/. Both Contrast 4 and Contrast 5 weigh mostly heavily the cues center of gravity and phoneme duration. Contrast 4 distinguishes fricatives and the phoneme /k/ from non-fricatives, but it is ambiguous for the phonemes /g/ and /l/. Contrast 5 groups /g/ and /l/ with non-fricatives. All the fricatives have the same representation except for /v/, but this might be because the only word type with /v/ in the coda is "of."

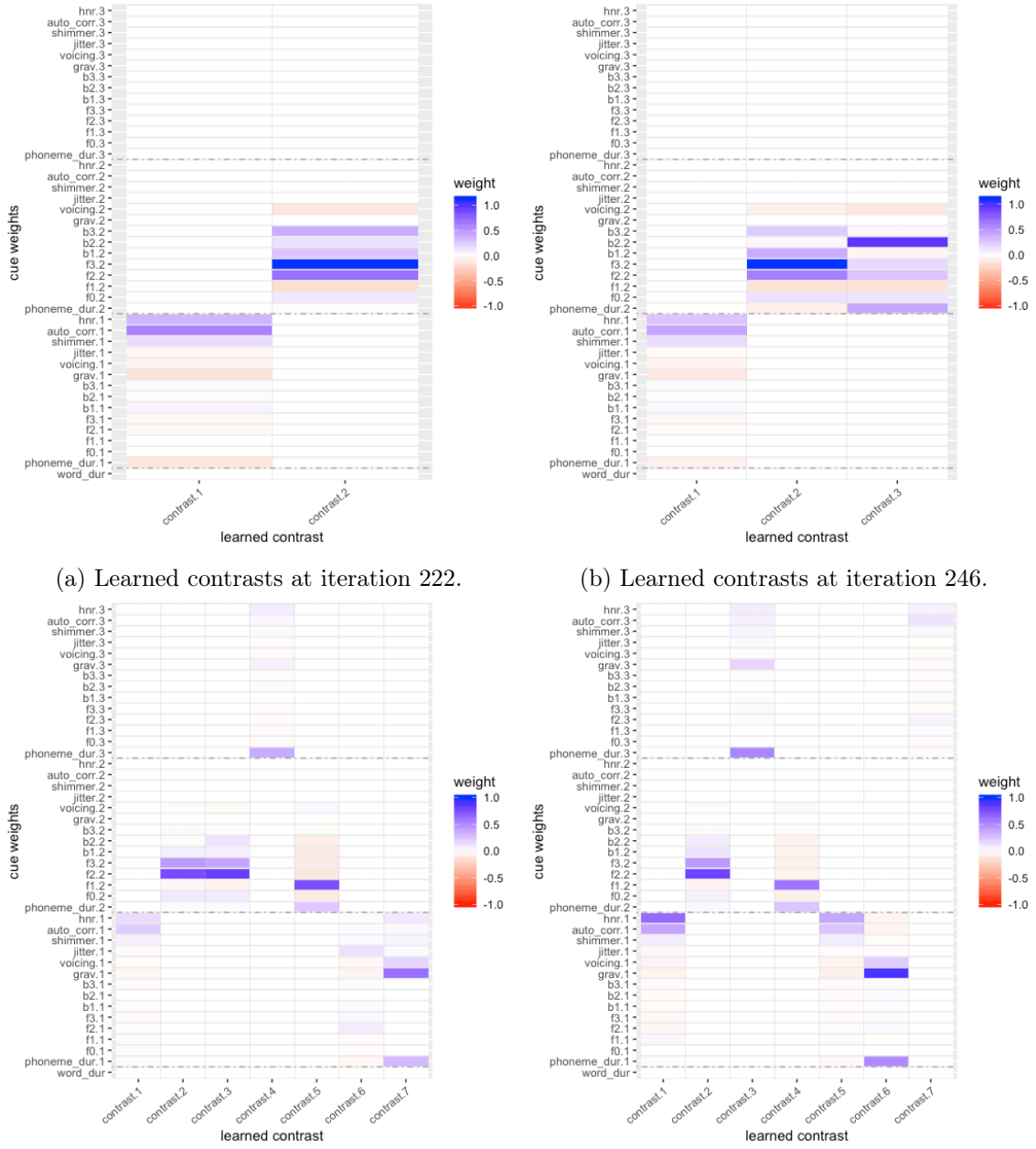
	Contrast 2 Voicing		Contrast 4 Fricatives + k		Contrast 5 Fricatives - v	
	0	1	0	1	0	1
∅	14.29	85.71	71.43	28.57	100.00	0.00
p	100.00	0.00	83.33	16.67	83.33	16.67
d	28.57	71.43	100.00	0.00	85.71	14.29
t	80.00	20.00	80.00	20.00	100.00	0.00
g	50.00	50.00	50.00	50.00	100.00	0.00
k	100.00	0.00	0.00	100.00	100.00	0.00
f	100.00	0.00	0.00	100.00	0.00	100.00
v	0.00	100.00	100.00	0.00	100.00	0.00
s	100.00	0.00	0.00	100.00	0.00	100.00
ʃ	100.00	0.00	0.00	100.00	0.00	100.00
θ	100.00	0.00	0.00	100.00	33.33	66.67
z	20.00	80.00	0.00	100.00	0.00	100.00
l	50.00	50.00	50.00	50.00	100.00	0.00
ɹ	0.00	100.00	66.67	33.33	100.00	0.00

Table 3.6: Percentages of each consonant phoneme assigned to each side of a learned phonological contrast.

3.3.2.4 An example of category consolidation

The learning mechanism outlined in Section 3.2.2.3 describes a scenario where two contrasts can become similar and consolidate without any changes to the system of lexical contrast. This section shows a specific example of how this process is played out during the course of learning by the model.

Figure 3.17 and Table 3.17 show the four snapshots of the learning process in a particular trial. On iteration 222 (3.17a), the model learns a vowel contrast (Contrast 2) from WE and BUT since Contrast 1 does not distinguish between them. On iteration 246, the model learns another vowel contrast (Contrast 3) from DO and BUT. By iteration 974, Contrast 2 and Contrast 3 have become fairly similar. On iteration 1400, CONTRAST 2 and Contrast 3 are consolidated into one category. When the contrasts are initially learned, the acoustics cues that were weighted the heaviest do not make much sense based on what we know about English phonetics. F3 for Contrast 2 and B2 (bandwidth of F2) for Contrast 3 are not the most important acoustic cues when it comes to vowel distinctions (Figure 3.17b). As more words are learned and classified, both contrasts update with the phonetics of newly



(a) Learned contrasts at iteration 222.

(b) Learned contrasts at iteration 246.

(c) Learned contrasts at iteration 974.

(d) Learned contrasts at iteration 1400.

Figure 3.17: An illustration of contrast generalization.

acquired words and begin to make more sense phonetically. F2, which corresponds to the frontness/backness of the vowel, becomes more heavily weighted for both contrasts and eventually the two become functionally the same and are consolidated.

	contrast							
referent	1	2	3	4	5	6	7	8
iteration 222								
WE	1	1						
BUT	1	0						
iteration 246								
WE	1	1	1					
BUT	1	0	0					
DO	1	0	1					
iteration 974								
WE	1	1	1	0	0	0	0	
BUT	0	0	0	0	1	0	0	
DO	1	0	1	0	0	1	0	
iteration 1400								
WE	1	×	1	0	0	1	0	0
BUT	0	×	0	0	1	0	0	0
DO	0	×	1	0	0	1	0	0

Table 3.7: Evolution of learned lexical representations.

3.4 Discussion

The model presented in this chapter makes several important contributions to the understanding of first language acquisition and phonological representation. The model succeeds in learning phonological contrasts appropriate for a given lexicon by identifying meaningful boundaries in the multidimensional acoustic space. These results demonstrate the efficacy of a nonparametric and unsupervised approach to modeling phonological acquisition and that phonological features are an emergent property from structuring the acoustic space to accommodate lexical contrast.

3.4.1 Computational approach

The model advances the computational study of phonological acquisition in a number of ways. First, this model is a general model of phonological acquisition. Many previous computational models of speech category acquisition focus on specific contrasts and use

cues relevant to those contrasts as input for learning, such as vowels (Vallabha et al., 2007; Feldman et al., 2013a; Dillon et al., 2013) and voicing (Toscano and McMurray, 2010). The model presented in this chapter is not limited to specific contrasts but aims to learn any contrast in a given lexicon.

Second, it is common practice to use artificially generated data as input (e.g., Vallabha et al., 2007; Toscano and McMurray, 2010; Feldman et al., 2013a). This model achieved reasonable results using natural acoustic measurements taken from a speech corpus. Moreover, the input consists of acoustic measurements from entire words. The acoustic representations used in this study better approximate the multidimensional and continuous nature of the speech signal a learner receives. Although the approach used in this model is not a perfect representation of continuous speech signal, it nevertheless is an important step forward towards more realistic input representation in acquisition modeling.

Third, the model is set up to more closely simulate the actual learning process of a child. This model also has the advantage of being nonparametric. In contrast, models that rely on statistical learning, such as Bayesian models (e.g., Feldman et al., 2013a), need parameter tuning to achieve the best results. Additionally, the learning is completely online. The learner hears the input one at a time and updates their phonological knowledge as needed at each iteration of learning, just as a child might as they are exposed to more and more linguistic input. In contrast, many existing models rely on batch learning. While these algorithms can be adapted to be online (e.g., Vallabha et al., 2007), their implementation are often parametric. Moreover, the learning in this model is unsupervised. It does not learn from target representations, but rather discovers both contrastive dimensions and appropriate phonological representations through learning. Acoustics and lexical contrast are sufficient for the learner to form appropriate abstract representations. All of these properties closely approximate the actual challenge faced by the learner.

Finally, the learning outcome from the experiment validates the learning mechanism described in the model. The model learns the appropriate numbers of phonological contrasts given the size of the input lexicon, and it also learns the appropriate phonetics for

each phonological contrast. Because phonological contrasts and lexical representations interact and update dynamically, this model can offer some explanations for the developmental trajectory of phonology. At the beginning of learning, the model had limited numbers of contrastive dimensions because only a few words need to be assigned abstract representations. However, with more input and sufficient word frequency, the model learns more distinct representations for different lexical items. This can in part explain why early lexical representation appears to be underspecified. With a small vocabulary, the learner does not need phonologically detailed representations because there are fewer word distinctions that need to be represented. The success of the model so far indicates representational pressures indeed play a role in phonological acquisition.

3.4.2 Theoretical implications

Phonological features are a useful tool of phonological analysis, but as reviewed in Section 2.2.2, assuming a universal set of innate features has a number of issues. The model presented in this chapter operationalizes the acquisition of emergent phonological features, and the experiment results indicate that the learning mechanism proposed in this chapter is computationally viable. One important theoretical advance from this model is that it outlines a concrete path from multidimensional acoustic input to abstract representation. Although many conceptual models of phonological acquisition incorporate lexical learning (e.g., Jusczyk, 1997; Dresher, 2004; Werker and Curtin, 2005), most of these models have not been implemented computationally and tested.

The learning is both phonetically and linguistically motivated, and the acoustic input and learned contrasts reflect the multidimensional nature of phonetic cues in production and perception. The hypothetical binary contrastive dimensions can offer insights into why phonological systems tend to be symmetrical. For example, if a contrastive dimension is created to distinguish vowel height for /i/ and /æ/, it is easy to extend the same contrast to vowels like /u/ and /ɑ/ since there are shared acoustic cues. Lastly, this model can capture the role of language experience. Depending on the input, the specific order of acquisition of

contrasts can differ, but the end result will converge to distinct phonological representation of all the lexical items when the critical number of lexical items has been acquired.

3.4.3 Future directions

There are a few aspects of the work that needs further development. First, it would be ideal if the model learns contrastive dimensions and cue weights that more consistently align with results from linguistic analysis. Although the results presented above are fairly close to linguist contrasts, the learning results vary from trial to trial. Part of this variation is expected, since there is a random element in word acquisition. However, the learning results might be more consistent with additional acoustic measurements. Second, at the maximum, only 210 lexical types were used as input to the model. It would be interesting to see how further input would alter the learning outcome of the model. Third, as this model is intended to be a general model of acquisition, the learning mechanism in the model should validated with results from additional languages. Lastly, this model only learns position-specific contrasts. Generalization across different positions is an important part of phonological learning and should be incorporated into a model phonological acquisition.

3.5 Conclusion

The learning model presented in this chapter makes several important contributions. First, it demonstrates that innate features are not necessary for the acquisition of discrete phonological representation. Second, it contributes to the research on emergent phonological features by proposing a clear mechanism whereby phonological contrasts can be learned from the input in a nonparametric and unsupervised fashion. Third, the model provides explanations for the trajectory of phonological acquisition observed in developmental studies. Overall, the results in the chapter suggest that phonological representations can emerge from the interaction of acoustics and lexical contrast without innate features or statistical learning.

Chapter 4

Lexical and Frequency Effects in Phonological Development

In this chapter, I provide developmental evidence that lexical contrast is a crucial cue in the acquisition of phonological distinctions. I use the Providence Corpus (Demuth et al., 2006) to 1) quantify the extent to which lexical contrast cues are present in the parental input, and 2) demonstrate that lexical contrast cues are predictive of phonological acquisition as measured by child production accuracy. The results show that minimal pair cues are abundant in both parental and child speech. Moreover, minimal pairs are predictive of production accuracy on both the word and the phoneme levels, indicating that the structure of lexical contrast plays an important role in phonological acquisition.

4.1 Background

The idea that phonological representation emerges from generalization over lexical items has been around for a long time. The early work by Ferguson and Farwell (1975) outlines a sketch of a phonological acquisition model where "children learn words from others, construct their own phonologies, and gradually develop phonological awareness." Their conception of acquisition emphasizes "the primacy of lexical items" and "individual variation." Subsequent work on phonological acquisition led to more fully developed conceptual models as well as growing experimental evidence on the interaction between the lexical development and phonological acquisition. There are many ways that lexical or sub-lexical cues can influence the formation of phonological categories, and a detailed study of child production can shed

light on the factors that play a role in phonological development.

4.1.1 Development of child production

Over the first year, the infants' vocalizations become increasingly speech-like. Before the onset of referential word use, infants often produce vocalic forms with stable meanings that do not correspond to any adult models, and these forms have been termed "protowords" or "quasi-words" by some researchers (Menn, 1976, 1983). Several studies suggest that protowords contain emergent phonological structures, and language-specific effects can be observed in early vocalizations (Menyuk et al., 1979; Stoel-Gammon and Cooper, 1984; Vihman and Miller, 1986).

For instance, adult listeners are able to discern the differences in vocal productions of 8- and 10-month-old infants acquiring French, Cantonese, and Arabic (de Boysson-Bardies et al., 1984). Additionally, differences in laryngeal articulation has been found in the babbling between infants acquiring English, Bai, and Arabic (Esling, 2012). The babbling vowel space is different for 10-month-old infants acquiring Algerian Arabic, Hong Kong Chinese, London English, and Parisian French (de Boysson-Bardies et al., 1989). The early babbling consonant repertoire appears to be similar across languages, with the majority of the consonants being stops, nasals, and the glide [h] (Locke, 1983), even when the ambient language does not contain /h/ as a phoneme (Vihman, 1992). However, there are significant cross-linguistic differences in consonant babbling that reflect the external linguistic input (de Boysson-Bardies and Vihman, 1991; Vihman et al., 1994). Moreover, Whalen et al. (1991) found significant differences in intonation patterns of reduplicated two- and three-syllable forms between English- and French-learning children between 6-12 months.

A general observation about infants' earliest words is that they tend to be surprisingly close to the adult targets. This observation has been ascribed to a process known as preselection, whereby infants "choose" to produce words that contain sounds which are more similar to their babbling repertoire and avoid words that contain more difficult sounds (e.g., Ferguson and Farwell, 1975; Fikkert and Levelt, 2008; Menn and Vihman, 2011). While this

avoidance of difficult sounds has been attributed to potential metalinguistic knowledge about the sounds themselves (Menn, 1983), Vihman (1991) argues for an alternative explanation which she terms the *articulatory filter*. In Vihman's (1991) account, the infants use both bottom-up and top-down knowledge in their learning and production of words. In other words, infants selectively produce certain sound sequences as the result of the interaction and reinforcement between their own familiar articulatory sequences and similar sequences in the input. There is experimental evidence supporting this view (Vihman et al., 2014).

After the accurate initial production, there is usually a drop in the similarity to the adult targets in the child's production. This kind of U-shaped development of linguistic ability has been observed in other domains of first language acquisition, especially in inflectional morphology (Cazden, 1968; Marcus et al., 1992). A child may initially correctly produce irregular forms (e.g., fall → fell), but as they acquire the -ed past tense rule, they will often overgeneralize this rule to irregular forms and produce "falled", leading to a drop in overall accuracy. The drop in word form production accuracy has similarly been attributed to increasing phonological systematicity (e.g., Macken and Ferguson, 1983; Vihman and Velleman, 2000).

4.1.2 Lexical and sub-lexical factors in phonological development

The development of performance accuracy in child production has been the subject of many previous studies. A number of probabilistic and distributional lexical and sub-lexical cues have been found to have some effect on the phonological and word learning. Other factors such as vocabulary size and phonological neighborhood density have also been put forward as explanations for phonological development. Nevertheless, the observation of an effect from any of these factors does not mean that these factors are necessary for the development of a linguistic system. It is, therefore, important to evaluate the relative contributions of these factors in phonological development.

4.1.2.1 Frequency

With the rise in popularity of distributional learning as an explanation for language acquisition, the role of frequency in the linguistic input has received considerable attention in acquisition studies (Ellis, 2002). Experimental and corpus studies have found frequency effects on phonemic and lexical acquisition.

On the phoneme level, the vowel space of infant babbling tends to reflect phoneme frequencies of the ambient language (de Boysson-Bardies and Vihman, 1991). For instance, English learning children's coda production has been found to match English coda frequencies (Zamuner et al., 2005). Also, Ingram (1988a) suggests that English-learning children usually acquire /v/ late because /v/ is not a frequent phoneme in English, while children learning Swedish, Estonian, and Bulgarian acquire this sound earlier because they are more frequent in the lexicon. Additionally, Beckman and Edwards (2010) shows correlation between phoneme frequency and consonant production accuracy for English- and Cantonese-learning children. Edwards and Beckman (2008) looked at production accuracy of word-initial consonants for 2- and 3-year-olds and concluded that language acquisition is influenced by both universal constraints and language specific frequencies.

Frequency effects have also been observed on the level of word learning. While the total frequency of words are not predictive of child production of these words, the frequency of words uttered in isolation in the input is predictive of child word production at a later date (Brent and Siskind, 2001). Furthermore, there is evidence that more frequent words tend to be learned more accurately. Japanese learning children aged 1;5–2;1 are less likely to truncate words that are more frequent in the maternal input (Ota, 2006). Additionally, frequency interacts with positional salience in predicting the child's production of lexical items. For Italian-learning children aged 1;4–1;8, the occurrence of nouns in utterance-final positions in the input predicted the production of nouns, while the occurrence of verbs in utterance-initial positions is correlated with verb production (Longobardi et al., 2015).

Overall, it appears that frequency has some effect on the acquisition of phoneme and word production. However, the effect is not always straightforward and the interaction with

other factors sometimes needs to be considered.

4.1.2.2 Phonotactic probability

In phonological analysis, phonotactics refers to the restrictions on the combinations of phonemes. For instance, /kn-, pt-, ps-, sr-/ are not permissible onset clusters for English, but /sp-, tr-, gl-/ are. An English speaker might judge a made-up word like /srum/ to not be a possible word of English, while /spum/ could be a word in English. Phonotactics is part of the speaker's implicit knowledge about the phonology of their language and part of their linguistic competence. Along with the rising interest in applying statistical learning to various acquisition problems, phonotactic probability has been proposed to play a role in lexical acquisition. Unlike phonotactics which describes patterns of possible and impossible sound sequences in discrete terms, phonotactic probability quantifies the likelihood of sound combinations through the frequencies of the co-occurrences of sound sequences in a language. Higher phonotactic ability may facilitate speech processing on the sublexical level (Vitevitch and Luce, 1998, 1999).

There is some evidence that phonotactic probability influences phonological and lexical acquisition, but the results are inconclusive. Infants show preference for sound sequences with higher phonotactic probability at 9 months (Jusczyk et al., 1994), and older children (aged 3;2-6;3) learn words with higher phonotactic probability with more ease (Storkel, 2001). Another study showed that children aged 3;2-8;10 can repeat non-words with frequent phoneme sequences with higher accuracy (Edwards et al., 2004). On the other hand, 4-year-olds are more accurate at learning words with rare sound sequences (Storkel and Lee, 2011). For older children, no effect of phonotactic probability was found in the learning of nonwords by 7-year-olds, while 10- and 13-year-old children had an easier time learning high probability non-words (Storkel and Rogers, 2000).

4.1.2.3 Vocabulary size

The effect of vocabulary size on word learning and phonological tasks is more consistent. Vocabulary size has been found to predict children's performance on word learning and phonological tasks. At 14 months, children with larger vocabulary find it easier to learn minimally contrasting non-words (Werker et al., 2002). For continuous speech processing at 18 and 21 month, children with larger productive vocabulary were more accurate and faster at responding to familiar words (Fernald et al., 2001). At ages 3-5, children with larger vocabulary tend to be more accurate at non-word repetition (Metsala, 1999). Although Edwards et al. (2004) found effects of phonotactic probability, children with larger vocabularies showed less frequency effects. In a subsequent study that included children with specific language impairment, Munson et al. (2005b) found that these children performed similarly as their vocabulary size matched peers, and overall vocabulary size is the best predictor of non-word repetition accuracy. These results suggest that greater vocabulary sizes provide the learner with the opportunity to generalize phonological contrasts over more words, resulting in better phonological awareness and word learning abilities.

4.1.2.4 Minimal pairs and related concepts

In traditional phonological analysis, minimal pairs are used to establish the phonemic inventory of a language. Minimal pairs refer to two words that are distinct in meaning and differ by one phonological unit. As this definition stands, phonological neighborhood density and functional load are very similar concepts, but they tend to be used in different contexts.

The phonological neighbors of a word is defined as the set of words that can be obtained by adding, subtracting, or substituting one segment of this word (Luce, 1986). Calculating the number of minimal pairs and the neighborhood density is methodologically similar. The main difference between them is that in practice, linguists tend to restrict minimal pair analysis to words of the same lengths. For instance, pairs like "cold" and "gold" can be used to establish the contrastiveness of /k/ and /g/, while pairs like "old" vs. "cold" and "old" vs "gold" are rarely used in such analyses even though both pairs technically differ by

one segment. Minimal pairs and phonological neighborhoods differ mostly in how they are used rather than how they are identified and calculated. Phonological neighborhoods are commonly used in models and experiments in speech processing, while phonologist and phoneticians use minimal pairs to identify and study properties of specific linguistic contrasts.

Another related concept is functional load, which tends to be used in work on sound change as a measure of the importance of a phonological contrast in the lexicon (Martinet, 1952; Wedel et al., 2013). If a phoneme is used to distinguish many words, it has a high functional load. Methodologically, a phoneme's functional load is often quantified as the number of minimal pairs it distinguishes. In a study on mergers, Wedel et al. (2013) shows that phonemes with lower functional load (i.e., fewer minimal pairs) are more likely to merge than high functional load phonemes.

Even though minimal pairs have been used in phonological analysis for a very long time, there has been few studies on first language acquisition that explicitly look at the interaction between minimal pairs and acquisition results. In second language acquisition, however, minimal pair training has been found to improve both perception and production of second-language phonemic contrasts (Logan et al., 1991; Bradlow et al., 1997; Wang et al., 1999), and minimal pair training results in better discrimination abilities than perceptual training alone (Hayes-Harb, 2007). However, given the differences between first and second language acquisition, these findings do not imply that minimal pairs are also predictive first language acquisition outcomes. The study in this chapter is intended to fill the gap in the general lack of direct study on the role of minimal pairs in first language acquisition.

4.1.3 Quantifying linguistic competence from linguistic performance

Since the primary evidence for this study comes from a corpus of child production data, it is necessary to carefully consider the relationship between linguistic performance and linguistic competence in drawing conclusions from such an analysis. Phonological competence, like any other level of linguistic knowledge, is part of the speaker's I-language, i.e., the internal mental representation of their language (Chomsky, 1986). The obvious challenge to the

study of I-language is that barring some exceptional advances in neuroliguistics, it cannot be directly observed. With mature speakers, it is possible to indirectly study the nature of their I-language experimentally or through linguistic tasks such as grammaticality judgments. However, studies aimed at understanding children's developing I-language are limited by practical concerns when working with infants and young children.

The limits of experimental data on early perception has been reviewed in the previous chapters. Essentially, the discrepancy between perceptual discrimination and word learning results show that perceptual discrimination should not be used as the sole evidence for the existence of phonological distinctions in the internal grammatical representation of the child. These studies nevertheless reveal something about the units of perception in early language learning, which are the necessary precursors for adult-like phonological units. There are two interpretations for the disparity between young children's phonetic and phonological performance: 1) Young children are phonologically competent but their performance suffers from non-linguistic factors like cognitive processing demands and motor control skills, and 2) young children have not developed complete phonological competence yet, and hence the poor performance.

If perceptual results offer limited but not conclusive indications of the state of development of a child's language, what about production? This chapter uses production accuracy from the Providence Corpus as a proxy of phonological competence. There are definite concerns with this approach since linguistic competence and linguistic performance are not equivalent, especially for children with developing motor control skills. Observations like the fish-phenomenon calls into question the validity of using child production to measure phonological knowledge. First documented by (Berko and Brown, 1960), the fish-phenomenon describes a situation in which a child misses pronounces a word [fɪs] for "fish", but rejects the pronunciation by an adult when it is repeated back to the child (Smith et al., 1973).

The wrong production of a word or phoneme can be the result of either linguistic performance or competence: It is possible that the child has not arrived at an adult-like representation or has trouble executing the specific sequence of articulatory gestures. However,

it would be unreasonable to attribute *consistently accurate* production of word forms and phonemes to mere performance. As discussed earlier, child production tends to follow a U-shaped curve in terms of target-like accuracy. Before reaching phonological competence, we should expect to see variation in production accuracy as the result of the systematicization of the phonological system. Sporadically accurate production is not informative about the child's linguistic competence, only that phonological reorganization is taking place. However, if the production data shows consistent accuracy towards the adult targets, it is possible to draw conclusions about the child's linguistic competence.

4.2 The Providence Corpus

The Providence Corpus consists of recordings and transcripts of spontaneous mother-child interactions for six monolingual English-learning children (Demuth et al., 2006). The recordings made by Katherine Demuth and her research assistants at the Child Language Lab at Brown University in Providence, RI. Data collection occurred between the years 2002-2005, with a total of 364 hours of recorded video and audio data. The recordings were carried out every two weeks and they usually occurred at the homes of the subjects. Each recording session was approximately one hour long. The Providence Corpus was chosen for this study because 1) it contains naturalistic data of both the parental input and the child production, 2) the children recorded in the corpus were in the age range (1-3 years) of interest for phonological acquisition, 3) there is sufficient data for each child, and 4) this corpus has been orthographically transcribed for both parents and children, and 5) phonetic transcription made by trained transcribers is available for all the children. A summary of the children from the corpus is provided in Table 4.1.

To conduct phonological analysis of the Providence Corpus, the existing transcription first needs to be processed into a form suitable for the goals of this study. Different types of transcriptions are available for the parents and the children. The parental speech has been transcribed with standard orthography and marked for part of speech, morphological stem, and position in the utterance. The corpus does not provide phonemic or phonetic transcrip-

Name	Age Range	Sessions	Sex
Alex	1;04.28-3;05.16	51	M
Ethan	0;11.04-2;11.01	50	M
Lily	1;01.02-4;00.02	80	F
Naima	0;11.27-3;10.10	88	F
Violet	1;02.00-3;11.24	51	F
William	1;04.12 - 3;04.18	44	M

Table 4.1: Summary of the information about the children and recordings in the Providence Corpus.

tions of the parental speech. The children's speech has been transcribed for orthography, actual produced phonetic forms, target phonemes, part of speech, morphological stem, and position in the utterance.

4.2.1 Processing of parental speech

For the parental speech, the orthography, phonemic transcription, morphological stem, and part of speech were obtained for analysis in this chapter. Because the parental speech was not transcribed phonemically in the corpus, it is necessary to first obtain phonemic transcriptions before any phonological analysis can occur. Phonemic transcriptions were applied to the orthographic forms of the parental speech from the CMU Pronouncing Dictionary, which covered the majority of the orthographic transcriptions in the corpus. There was a number of frequent words whose phonemic transcriptions were not available from the CMU Dictionary. For relatively more frequent words (>20 occurrences in the entire corpus) such as content words (e.g., lollie, scrumptious, hummus), diminutive forms (blankie, nursie, piggie), proper names (Naima, Eeyore, Mufasa), an additional dictionary was created where the phonemic transcriptions were manually entered. Other words were excluded from this analysis. The excluded words include unintelligible speech (e.g., xxx, www), interjections (e.g., uhoh, uhhuh, tadah), highly reduced forms (e.g., dya, whaddya), and less frequent words (<20 occurrences in the entire corpus). In total, 4911 word types were excluded. Even though this seems to be a large number, most of the words were of the types described above, and 2739 only occurred once in the entire corpus. The transcriptions were converted from Arpabet in the CMU dictionary to the IPA so that it easier to compare with the IPA

transcriptions of child production.

Stem-level transcription is available from the corpus. For example, a word with the orthographic representation of "hats" would be transcribed on the stem level as "hat-pl". From stem-level transcription, the root of each word was found by removing the suffixes. Thus, for "hat-pl", the root is simply "hat". Additionally, the corpus transcription has detailed part of speech (POS) tags. For example, the pronoun category "pro" is further divided into subcategories, such as demonstratives ("pro:dem"), relative pronouns ("pro:rel"), and indefinite pronouns ("pro:indef"). From these labels, the larger, more basic POS categories (e.g., just "pro") were obtained for each word.

After each word was processed, exclusions of certain words were applied as follows. Words without POS tags were excluded, and these were almost exclusively space fillers (e.g., um, uh, ooh, aw). Some words labeled with the POS tag "co" (communicator) were also excluded; these include fillers such as "mhm", "huh", "wah". Communicator words like "yeah", "okay", and "please" were kept. Further exclusions based of POS tags include "sing" (8 tokens for when the parent was singing), "none" (2 tokens), "chi" (120 tokens of child-invented forms like "dede", "wa", "balog"), "wplay" (157 tokens, e.g., "phooey", "snip"), "neo" (10 tokens, e.g., "tso", "skinks").

4.2.2 Processing of child production

For each word in the child speech, the child's actual production, target phonemes, morphological stem, and POS categories were obtained directly from the corpus. The transcriptions of each child's actual productions and the target forms were available from the corpus. Words were excluded if they did not include a target or actual transcription from the corpus. Most of the exclusions were unintelligible forms (e.g., xxx, yyy), and only 191 word types were excluded in total. Out of the 191 words, 119 only occurred once, and 175 had frequencies of 5 or less. Although both target phonemic forms and actual phonetic productions were transcribed in the IPA, further processing of the transcriptions was carried out to eliminate internal inconsistencies. For example, /ɚ/ was transcribed in a number of different ways

(e.g., ʌɪ, əɪ, ər, ɝ), and these were all standardized to /ɚ/, and affricates were transcribed both as digraphs (e.g., dʒ, tʃ) and as single letters (e.g., ʤ, ʧ), and these were consolidated as single letters. Moreover, diphthongs were represented as single units in the analysis conducted in this chapter.

After the processing of phonemic and phonetic transcriptions, the morphological stem of each word and the basic POS categories were derived in the same way as for the parental speech. Similar to the adult speech, words without POS markers were excluded. These were mostly unintelligible babbling transcribed as "xxx" or "yyy". Several other POS categories excluded from analysis include "fam" (with only one word "zoob"), "L" (116 tokens with no orthographic transcription at all), "neo" (15 tokens of nonwords like "vrap", "tso"), "cm" (557 tokens of words such as "uh", "um", "sssh"), "wplay" (306 tokens, e.g. "zub", "pommy"). Words with "*" marked as the model production or "*" as the actual production were also excluded, and these were for the most part fillers like "um" or sounds like "ss", "wa". An additional 1,692 words excluded are words whose orthographic representations are not found in parental speech. Most of these are low frequency forms. About half of these words (814) only have one occurrence in the entire corpus. These words include more communicator type words like "uhuh", "uhhuh", "tadah", and "mkay". Some of these other words in this excluded group demonstrate overgeneralization by the child, like "falled", and some of the words are the result of the transcriber attempting to transcribe phonetically with orthography; for example, data from one child had "goldipocks", "goldidocks", "goldisocks", and "goldiblocks", each with frequency of 1 or 2.

Some words were missing orthographic representations, but the transcription for stem, model production, and actual production were all available. There are 662 of these items, and most of them (judging from the phonetic transcriptions) are reduced forms of common words, such as "about", "around", "because". Also, in a few cases, the orthography appears to be a phonetic transcription. For instance, the word whose orthography transcribed as "ta" has the stem "to" and the POS "inf", but its actual production was marked as [tɑ]. Some orthographic transcriptions are misspelled, like "gree" for "green". Since most of these words

do not decompose further, the stem is used as a substitute for orthography.

4.2.3 Descriptive statistics of the processed data

Table 4.2 summarizes the post-processing data used for analysis in the rest of this chapter, including the total word counts for each of the participants, average word counts per session, the total number of unique orthographic words, and the number of unique stems. There appears to be individual variation in how much each parent and child talked, but some of the difference comes from the fact that some children were recorded until an older age. Naima and Lily have more total sessions because they were recorded weekly rather than every other week. Recordings of Lily, Naima, and Violet were carried out monthly between 3-4 years of age, while the other children were only recorded until they were around 3 years old.

Mothers					
child	sessions	word count	words per session	unique words	unique stems
Alex	51	144518	2833.69	3876	2555
Ethan	50	158607	3172.14	4518	2754
Naima	88	301420	3425.23	6240	3821
Lily	80	340372	4254.65	8370	5111
Violet	51	125525	2461.27	5265	3363
William	42	127042	3024.81	3539	2339
Children					
child	sessions	word count	words per session	unique words	unique stems
Alex	51	40102	786.31	1690	1279
Ethan	50	32057	641.14	2355	1705
Naima	88	112460	1277.95	3745	2446
Lily	79	77064	975.49	3081	2143
Violet	48	29226	608.88	1945	1403
William	44	34201	777.30	1658	1228

Table 4.2: Descriptive statistics of the data used for the analysis in this chapter.

4.3 Quantifying minimal pair cues in first language acquisition

This section of this chapter has a straightforward goal: to quantify the amount of minimal pair cues that exist in child-directed speech as well as child speech. To do so, I provide

minimal pair counts with different word exclusion criteria, and I also quantify the amount of minimal pair cues per session and between pairs of phonemes. This section of the chapter is meant to be purely descriptive, and the implications of minimal pair cues will be elaborated in the following section and the general discussion.

4.3.1 Methods

Although the definition of a minimal pair is quite straightforward, in practice, counting minimal pairs is a little bit more complicated especially in the context of language acquisition. The most straightforward examples of minimal pairs are words like "bad" vs. "bed", which differ by one vowel (/bæd/ vs. /bɛd/), and "pat" vs. "bat", which differ by one consonant (/pæt/ vs. /bæt/). However, in conversational speech, many words are inflected. Words like "hide" vs. "hid" differ by one phoneme, and their meanings differ with respect to tense. Being able to use "hide" and "hid" to learn the /ɑɪ/ vs. /ɪ/ distinction requires the child to have some knowledge that tense is a dimension of meaning difference, and young children may not have acquired this distinction early on. Additionally, the meanings of functional words are often rather abstract. For example, is it possible for the child to learn the /ð/ vs. /f/ distinction from "that" vs. "fat"? Additionally, the phoneme /ð/ rarely occurs in content words but is highly frequent in functional words. On the other hand, the abstractness of function words may not be a huge hurdle for learning phonemic categories from them since function words tend to occur in very different syntactic contexts than content words. If the child notices the word "that" consistently occurs in different positions than "fat", perhaps this distinction alone will enable the child to know that there is some difference between "that" vs. "fat".

In order to provide a full picture of minimal pair cues in parent-child interactions, minimal pairs were counted with different degrees of word exclusion. The word exclusions were meant to account for different scenarios where child may or may not be able to access certain word categories for phonological learning. In addition, phonological neighbors are also counted for a comparison with minimal pair measures. I present data from:

1. All the transcribed words from the processed corpus
2. Content words only (nouns, adjectives, verbs, and adverbs)
3. Monomorphemic words (content and functional)
4. Monomorphemic content words (nouns, adjectives, verbs, and adverbs) only
5. Frequent monomorphemic content words ($n > 10$)
6. Phonological neighbors of monomorphemic content words

To count minimal pairs, pairwise string comparisons were carried out between unique phonemic transcription types for all the parents, as well as unique target types for all the children. Two words were determined to be a minimal pair if they were the same length and differed by only one phoneme. For phonological neighbors, two words that differed in length by one are also included in addition to equal-length words. Two words were determined to be phonological neighbors if they differ by one phoneme through substitution, deletion, or addition. Phonological neighbors were only calculated for the monomorphemic content subset of the words.

Like phonological neighbor calculations, the subsequent analysis all used the restrictive monomorphemic content words to quantify minimal pair cues. This set of words were chosen to provide conservative estimates of what lexical contrast cues the child could use. Two measures computed from monomorphemic content words include the average number of minimal pairs each individual child heard and produced in each one hour recording session, as well as the numbers of minimal pairs for each pair of phonemes, both in parental speech and in child production.

4.3.2 Results

4.3.2.1 The effect of word exclusion criteria and method of counting

Figure 4.1 plots the overall numbers of minimal pairs counts for each phoneme according to different exclusion criteria delineated above. Unsurprisingly, the number of unique mini-

mal pair counts decreased as the word exclusion criteria became more and more restrictive. When all word forms were included in minimal pair count, parental speech included 32,648 unique minimal pairs in total, and the children produced 14,077 minimal pairs all combined. The total numbers decreased as word exclusions were applied. Counting only content words, the parents produced 27,919 minimal pairs, and the children produced 11,364. When only monomorphemic words were counted, parental speech had 16,191 minimal pairs, and the children had 8455. When only monomorphemic content words were considered, the counts were 9416 for the parents and 5059 for the children. For more frequent ($n > 10$) monomorphemic content words, the numbers decreased to 3647 words for the parents and 1631 words for the children.

While excluding words by type (content vs. functional), morphological complexity, and frequency reduced the number of minimal pairs, the relative numbers of minimal pairs between phonemes remain roughly the same. For instance, /d/, /k/, /t/ have more minimal pairs relative to other phonemes when all words were used to count minimal pairs, and this trend remained when functional words were excluded, when only root forms were used, when only root forms of content words were used, and when a frequency threshold was applied to the root forms of content words. Phonological neighborhood counts yielded likewise similar results overall; all the counts are slightly above monomorphemic content counts. The major difference is that phonological neighborhood counts included the correspondence of phonemes to null elements.

The observation that different exclusion criteria and counting methods result in similar trends is confirmed by pairwise correlations between the different minimal pair counts and the phonological neighbors count. Table 4.3 is a correlation matrix of all six measures of minimal pair for both the parents and the children, and it shows the correlation is high between different word exclusion conditions, minimal pair and phonological neighborhood counts, and parental and child counts. Because of this trend, different ways of minimal pair counting should have similar predictive power, as long as the method of counting and word type exclusion is systematic.

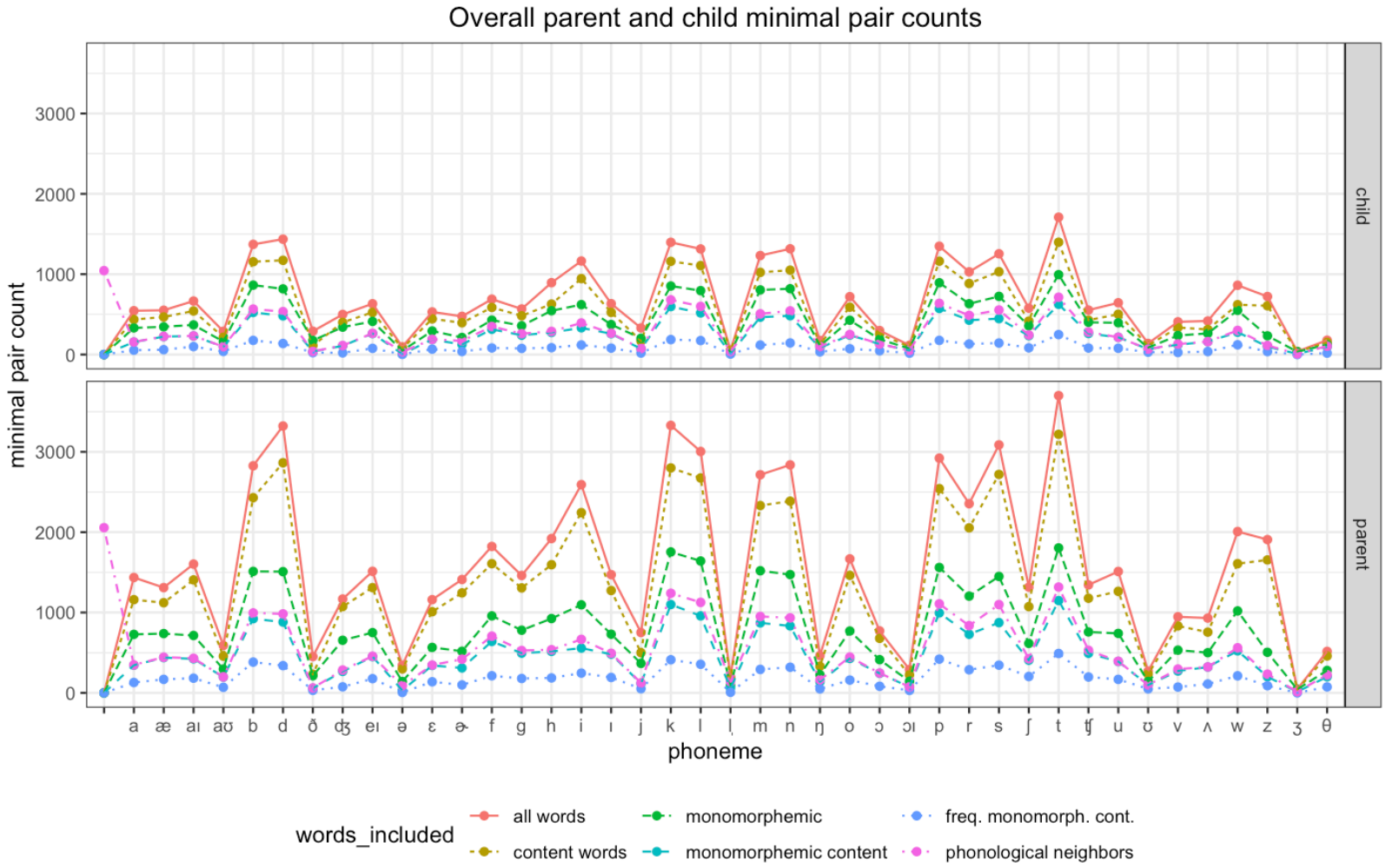


Figure 4.1: Comparison of minimal pair counts with different exclusion criteria.

	P-A	P-C	P-M	P-MC	P-FMC	P-PN	C-A	C-C	C-M	C-MC	C-FMC	C-PN
P-A	1.00	1.00	0.98	0.96	0.96	0.96	0.99	0.99	0.97	0.95	0.94	0.96
P-C	1.00	1.00	0.98	0.96	0.95	0.96	0.99	0.99	0.97	0.95	0.93	0.96
P-M	0.98	0.98	1.00	0.99	0.97	0.97	0.98	0.98	0.99	0.98	0.95	0.98
P-MC	0.96	0.96	0.99	1.00	0.99	0.99	0.96	0.97	0.98	0.99	0.96	0.99
P-FMC	0.96	0.95	0.97	0.99	1.00	1.00	0.96	0.97	0.97	0.99	0.98	0.99
P-PN	0.96	0.96	0.97	0.99	1.00	1.00	0.96	0.97	0.97	0.99	0.98	0.99
C-A	0.99	0.99	0.98	0.96	0.96	0.96	1.00	1.00	0.99	0.96	0.95	0.96
C-C	0.99	0.99	0.98	0.97	0.97	0.97	1.00	1.00	0.98	0.97	0.95	0.97
C-M	0.97	0.97	0.99	0.98	0.97	0.97	0.99	0.98	1.00	0.98	0.96	0.98
C-MC	0.95	0.95	0.98	0.99	0.99	0.99	0.96	0.97	0.98	1.00	0.97	1.00
C-FMC	0.94	0.93	0.95	0.96	0.98	0.98	0.95	0.95	0.96	0.97	1.00	0.97
C-PN	0.96	0.96	0.98	0.99	0.99	0.99	0.96	0.97	0.98	1.00	0.97	1.00

Table 4.3: Correlations between various minimal count measures and phonological neighbor counts for all the phonemes. Labels are abbreviated for space: "P-" = parental counts, and "C-" = child counts. A = all words, C = content words only, M = monomorphemic words, MC = monomorphemic content words, FMC = frequent monomorphemic content words, PN = phonological neighbors.

4.3.2.2 Minimal pair cues in natural speech

	All Words				Monomorphemic Content			
	Parents		Children		Parents		Children	
child	mean	sd	mean	sd	mean	sd	mean	sd
Alex	504.06	117.70	95.70	81.84	119.49	34.04	26.17	19.37
Ethan	578.24	205.11	121.28	76.09	145.52	51.36	29.57	17.09
Lily	810.08	204.81	177.13	101.32	222.69	77.21	38.72	22.63
Naima	545.92	139.15	168.40	105.79	141.05	46.78	37.80	24.75
Violet	471.75	174.33	116.76	69.02	120.06	64.59	22.12	15.50
William	483.71	228.76	112.30	99.90	110.98	59.99	21.78	23.18

Table 4.4: Means and standard deviations of minimal pair counts for the parents and children.

The above analysis counted minimal pair cues for each participant in the corpus across all their speech data. However, in each hourly recording session, is the child likely to hear many words that are minimal pairs? To answer this question, Individual analysis was carried out for both the parents and children for each session, and the numbers of minimal pairs were quantified per session. Table 4.4 shows the average counts per session using all the words and when only monomorphemic content words were considered. The distributions of minimal pair counts for monomorphemic content words for all the sessions are plotted in Figure 4.2. It appears that minimal pairs are a common occurrence in natural speech. Of

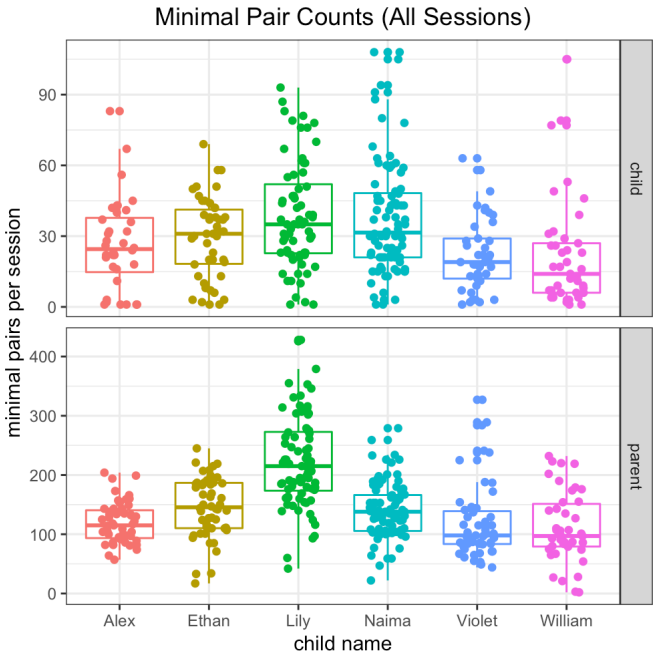
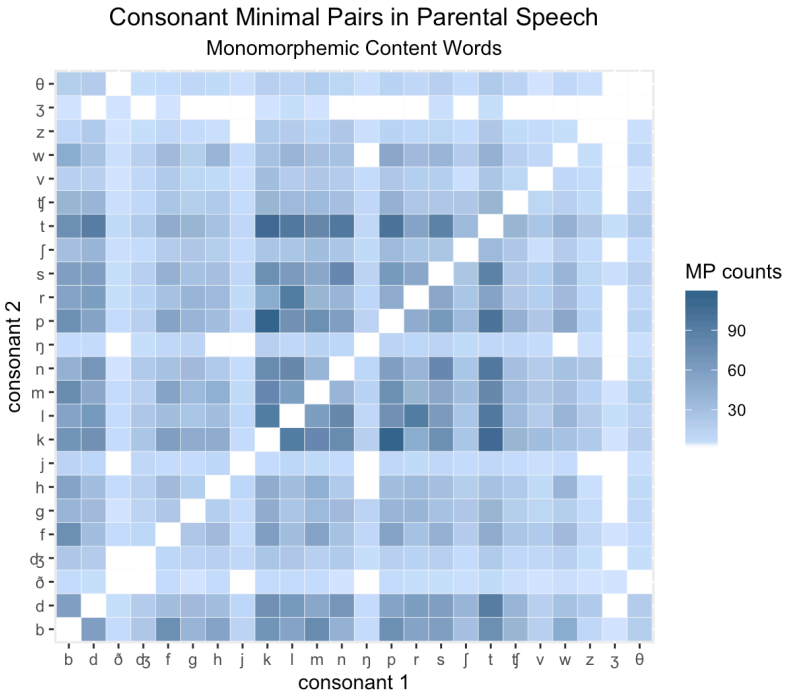


Figure 4.2: Boxplot of the number of minimal pairs in parental and child production for all the sessions.

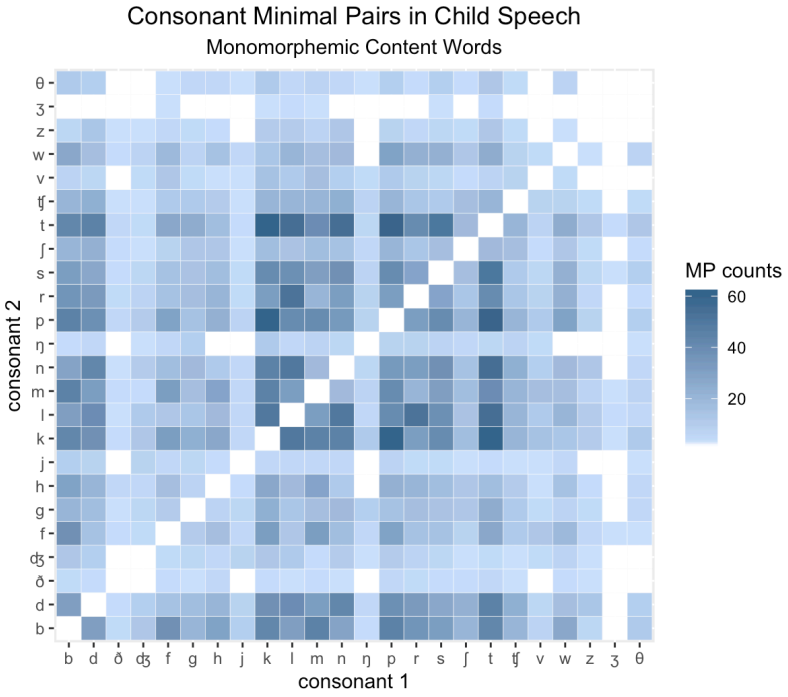
course, inclusion of more words results in high counts, but even when only monomorphemic words were considered, there were well overall 100 minimal pairs per hour of parental speech. It remains a question how much of this information the child can use in learning, but so far, the existence of minimal pair cues has been well demonstrated.

4.3.2.3 Minimal pair cues for pairs of phonemes

Although it is clear that minimal pair cues are abundant in natural speech, the question remains whether minimal contrast between phonemes can be learned from these minimal pairs. I present the results on minimal pair counts for each pairs of phonemes in parental and child speech. I include here only measures from the most restricted data set – the counts from only monomorphemic content words. Figure 4.3a and Figure 4.3b are heatmaps visualizing minimal pair counts for each pair of consonant phonemes for parents and children respectively. Most pairs of consonant phonemes are well represented. For both parents and children, the phoneme /ʒ/ has the fewest number of minimal pairs with other consonants. This is expected since /ʒ/ appears in relatively few words and typically restricted to word-

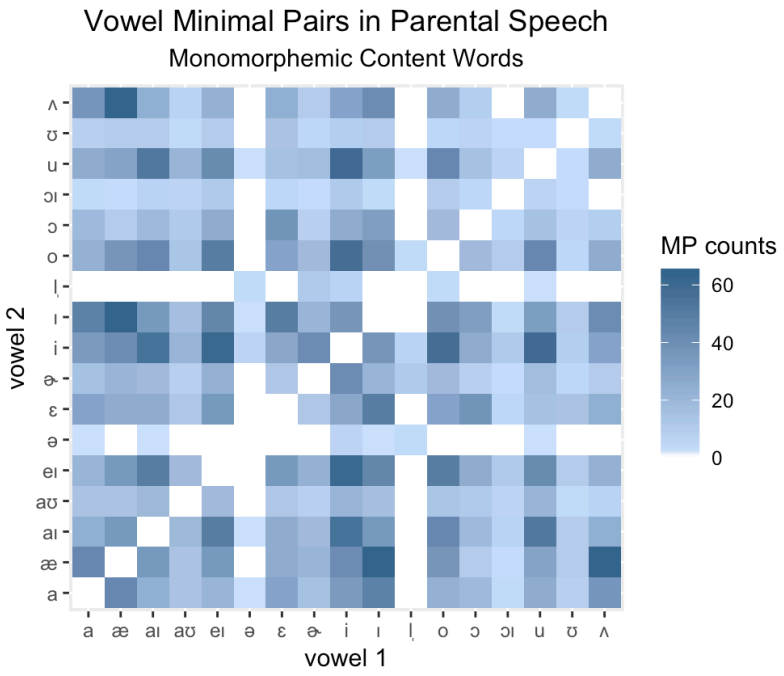


(a) Consonant minimal pair counts in parental speech.

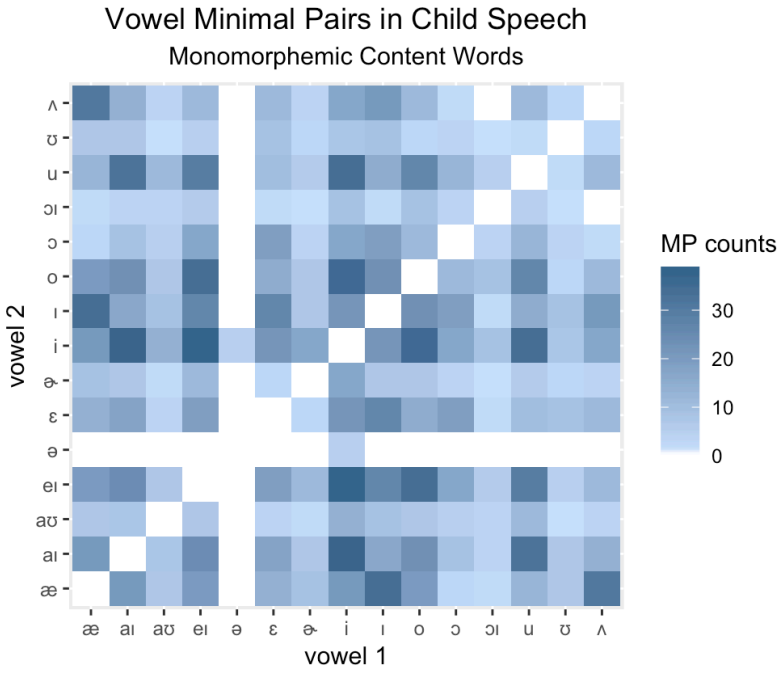


(b) Consonant minimal pair counts in child speech.

Figure 4.3: Unique minimal pair counts for each pair of consonant phonemes from both parental and child speech for monomorphemic content words.



(a) Vowel minimal pair counts in parental speech.



(b) Vowel minimal pair counts in child speech.

Figure 4.4: Unique minimal pair counts for each pair of vowel phonemes from both parental and child speech for monomorphemic content words.

medial contexts. This should not be problematic for acquisition, since the existence of any minimal pair with /ʒ/ should be enough evidence that a contrast needs to be learned. Similarly, vowels are well represented by minimal pairs in parent (Figure 4.4a) and child speech (Figure 4.4b). Of course, when more word categories are included, the minimal pair counts increase for all pairs of consonants and vowels.

4.3.2.4 Word frequency and minimal pairs

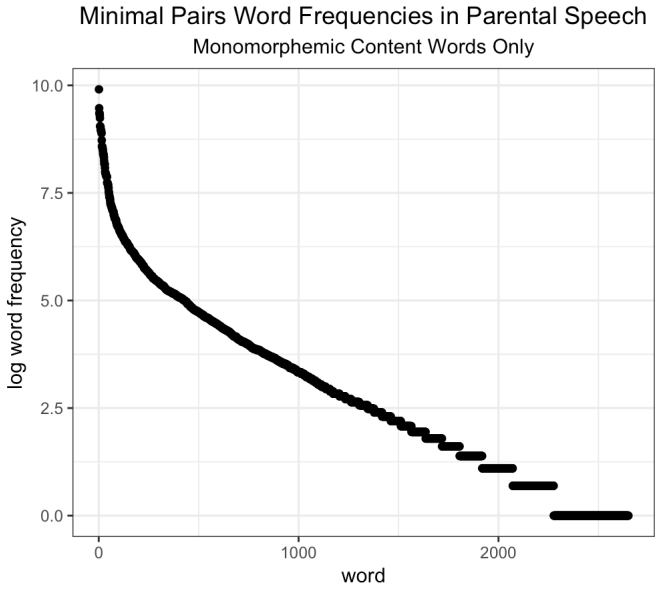


Figure 4.5: The frequencies of words included in the minimal pair count for parental speech. Only monomorphemic content words are included in this frequency count.

So far, I have demonstrated the abundance of minimal pair cues in parental speech and in child production. Nevertheless, the acquisition of phonological categories based on minimal pairs might also be influenced by how frequent these words occur. If the minimal pair of a highly frequent word seldom occurs, the child is likely to acquire the less frequent word and make use of the minimal pair contrast for phonological learning. The log word frequency is plotted in Figure 4.5. Out of 2650 unique words used in the minimal count, 374 occurred only once in the corpus. These words are not necessarily rare words in English, but they are perhaps less common in child-direct speech. Some examples of words that occurred only once include “mulch”, “caution”, “elite”, “feeble”, and the like. However, a large number

of minimal pairs are highly frequent. Of the monomorphemic content words from parental speech, 1297 pairs are words that have frequency counts above 50, from which 494 unique phonemic contrasts are represented.

4.4 An evaluation of factors in phonological acquisition

In the above section, I have demonstrated that minimal pairs are abundant in parental and child speech. However, the existence of minimal pair cues does not imply that the child can make use of it in learning phonology. In this section, I provide evidence that minimal pairs have an effect on phonological acquisition. I quantify child production accuracy on the word and phoneme levels and compare the effectiveness of minimal pairs, frequency, and phonotactic probability in predicting production accuracy.

4.4.1 Word level production

This section investigates word level production accuracy and which factors are predictive of word production accuracy. Production accuracy was quantified for each child, and the predictive power of word length, minimal pair counts, and word frequency were examined.

4.4.1.1 Methods

To quantify production accuracy for each child on the word level, the phonetic transcription of the child's actual production was compared to the target phonemic forms of each word. Two accuracy measures were calculated: categorical and gradient. For categorical accuracy, if the actual form is different from the target form by any segment, the production of the word is marked inaccurate; if the actual production matched the target form exactly, the word was marked as accurate. For gradient accuracy, the number of correctly produced segments is divided by the total number of segments of a word. For example, if a child produces [d.ʌ.d.i] for the target form "doggie" /d.ɑ.g.i/, this production would be categorically inaccurate but have a gradient accuracy of $2/4 = 0.5$. The overall accuracy of each word is calculated as the average accuracy over all productions of this word for both categorical accuracy and

gradient accuracy.

Minimal pair counts from the previous section are used in visualizations and statistical modeling in this section. In addition, orthographic word frequencies are calculated for each parent and child. Biphone phonotactic probabilities are calculated using Phonological CorpusTools (Hall et al., 2017) with log token frequencies as in the algorithm originally outlined in Vitevitch and Luce (2004). Linear regressions were estimated to test the effects of word length, minimal pair counts, word frequency, and phonotactic probability on production accuracy on the word level (word accuracy∼word length + minimal pairs + phonotactic probability + frequency). Because each measure was independently calculated for each child, separate linear regression models are estimate created for each child.

4.4.1.2 Results

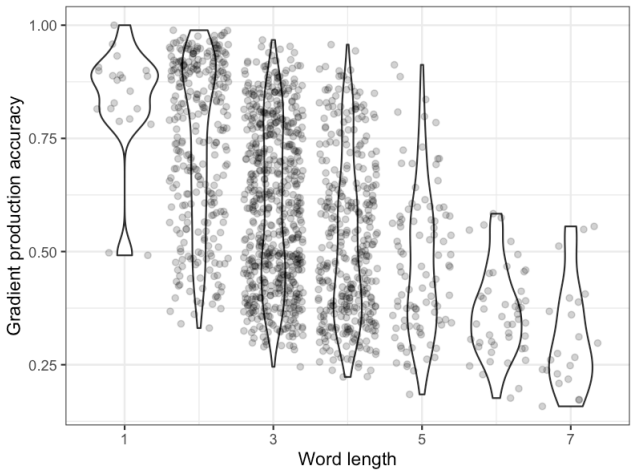
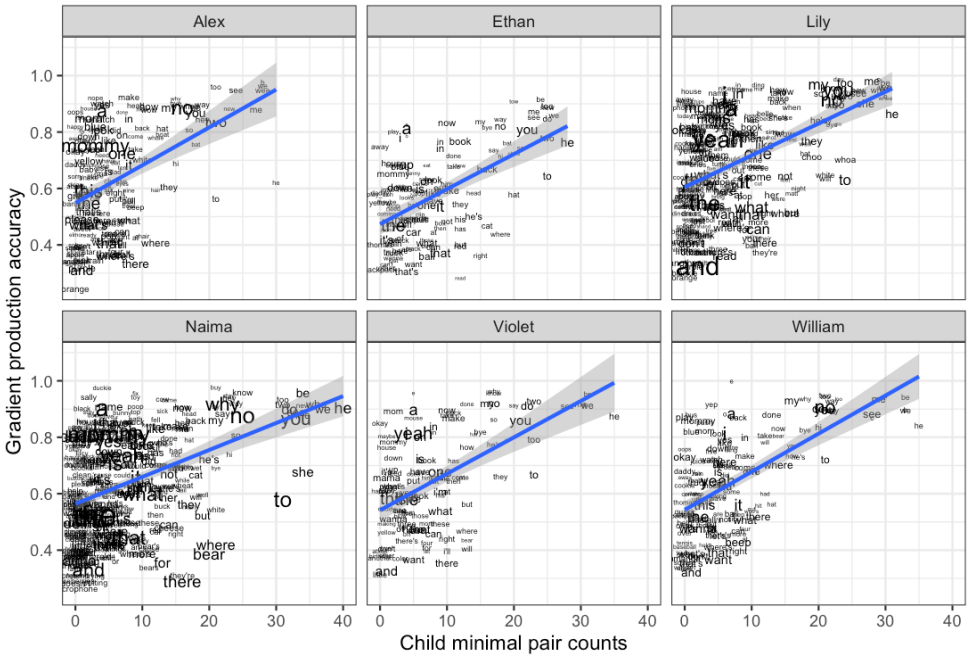
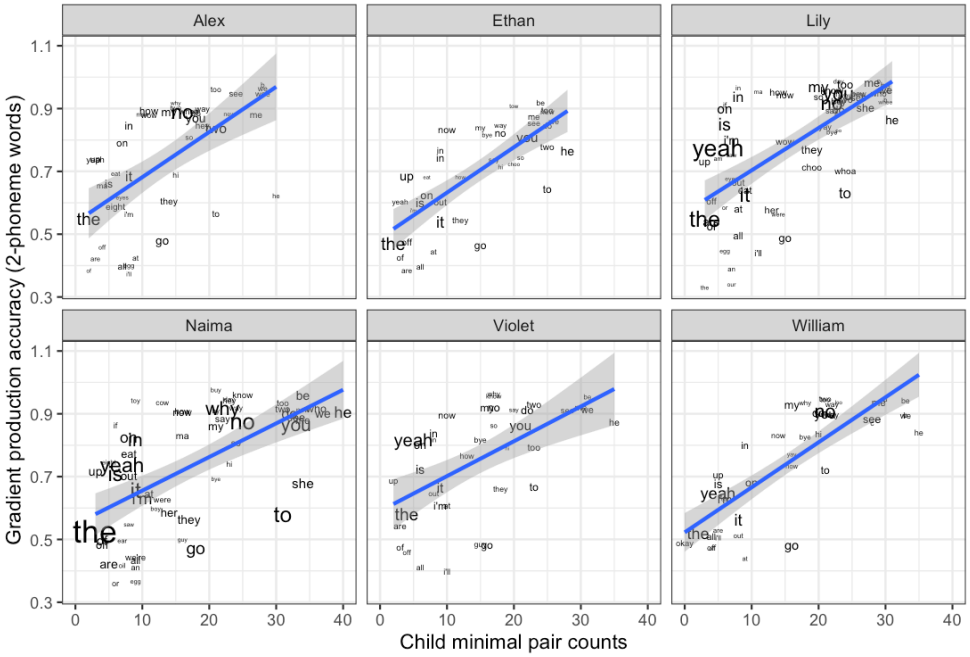


Figure 4.6: Word length and gradient production accuracy. Shorter words tend to be produced more accurately.

Word length. Figure 4.6 shows the effect of word length on production accuracy. Word length is quantified by the number of phonemes in the target forms of each word, and production accuracy is pooled from all six children. Unsurprisingly, shorter words tend to be more accurately produced than longer words. There is a wide range of production accuracy for all lengths of words, but on average, accuracy drops as word length becomes



(a) Child minimal pair counts and gradient word production accuracy for the six children for all words. There is an overall trend that more minimal pairs indicate better production accuracy.



(b) Child minimal pair counts and gradient word production accuracy for the six children for all 2-phoneme words. The trend that more minimal pairs indicate better production accuracy remains.

Figure 4.7: Minimal pairs and word production accuracy.

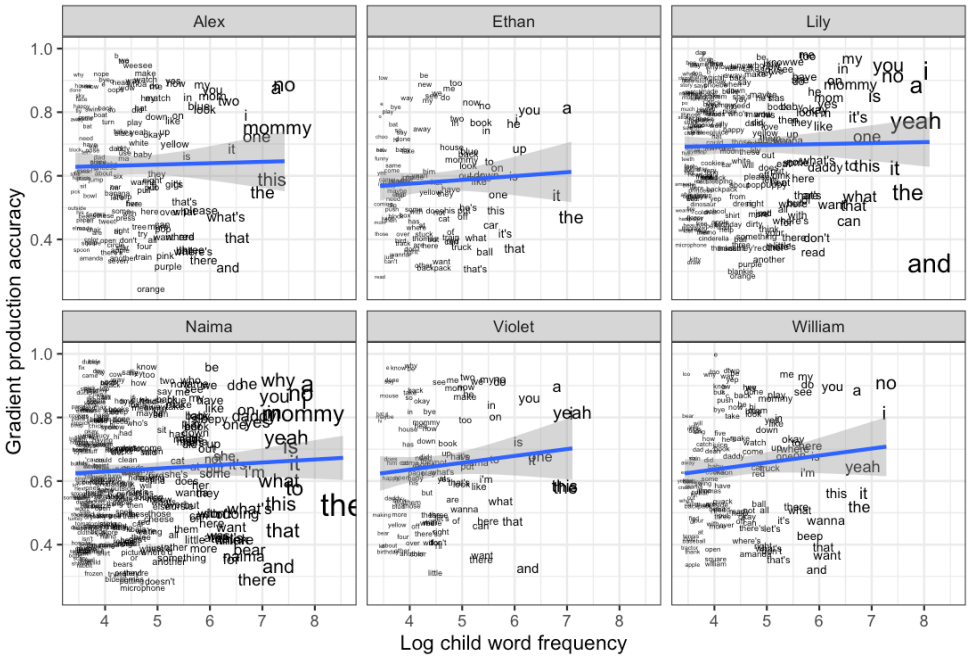
longer.

Minimal pairs. In addition to word length, the number of minimal pairs have an impact of production accuracy on the word level. The minimal pair counts used in this section are specific to each child rather than pooled across all the children. Figure 4.7a shows gradient word production accuracy[1] and child minimal pair counts for all the words in each of the children's production. There is generally an upward trend for all of them. The words clustered around the upper right corner (i.e., more accurately produced words) appear to be shorter. It is possible that these trends are the result of word length differences than minimal pair differences. To further evaluate the effect of minimal pairs on word production, 2-phoneme words for each of the children were plotted them themselves in Figure 4.7b. When limited to 2-phoneme words, the same trend is observed: Words with more minimal pairs are more accurately produced.

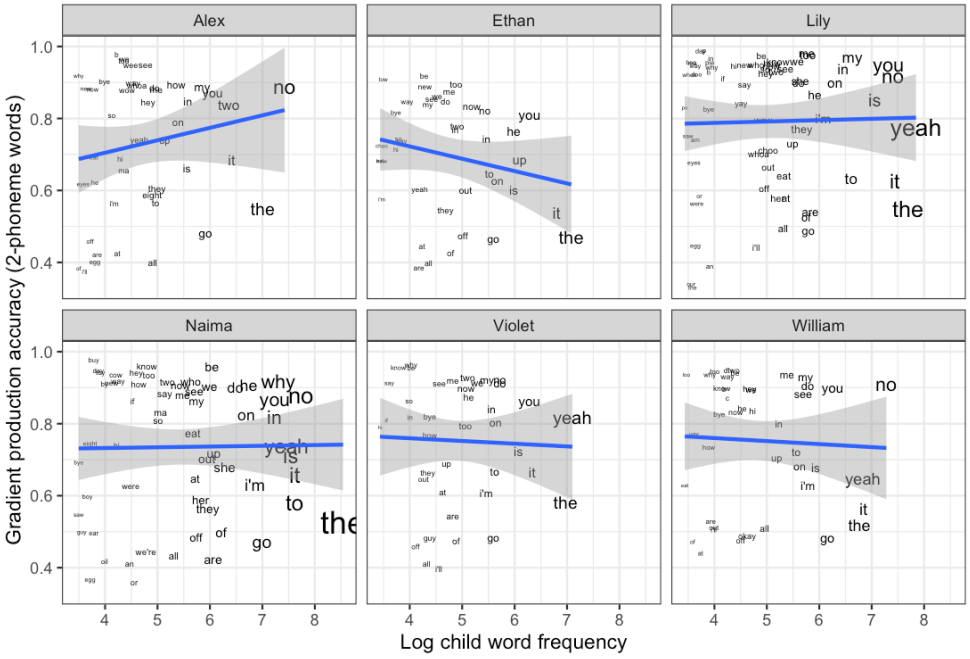
Word frequency. On the other hand, the same trend does not occur for word frequency. Figure 4.8a and Figure 4.8b plots word accuracy against child specific word-frequency counts for all the words and for 2-phoneme words respectively. Unlike the clear trends observed for minimal pairs, there are no patterns between word frequency and word production accuracy.

Phonotactic probability. Interestingly, there is to be a negative trend for phonotactic probability between word production accuracy and phonotactic probability. The trend appeared to be heavily affected by outliers for some of the children, but nevertheless it is a consistent pattern for all six children.

[1] Categorical word accuracy shows very similar trends. See Appendix A for a brief discussion.

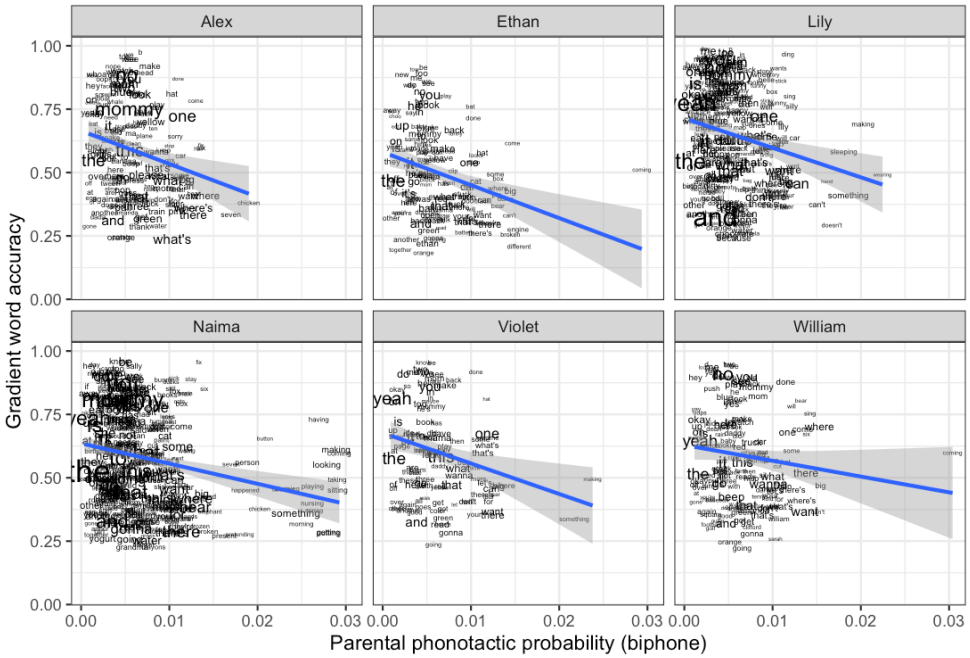


(a) Child word frequency and production accuracy for all words. There is no trend that more frequent words are more accurately produced.

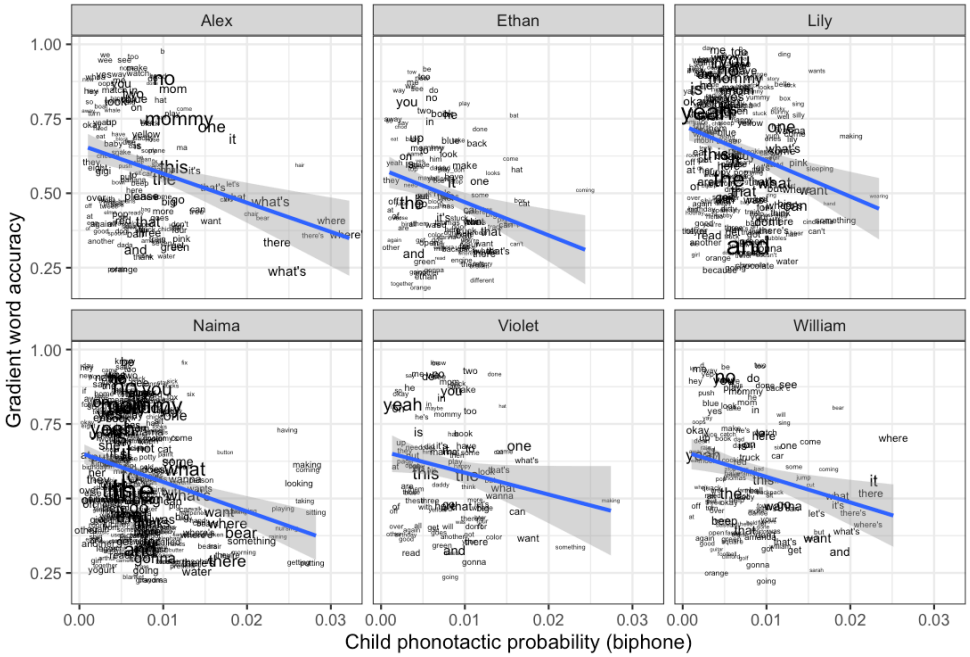


(b) Child word frequency and word production accuracy for the six children for 2-phoneme words.

Figure 4.8: Word frequency and word production accuracy.



(a) Parental phonotactic probabilities and child production accuracy for all words.



(b) Child phonotactic probability and production accuracy for all words.

Figure 4.9: Parental and child phonotactic probability and word production accuracy.

	Estimate	Std. Error	t value	Pr(> |t|)
		Alex		
(Intercept)	0.7268	0.0615	11.81	0.0000***
word length	-0.0448	0.0168	-2.67	0.0084**
child minimal pairs	0.0106	0.0023	4.51	0.0000***
child phonotactic probability	-8.6570	2.3533	-3.68	0.0003***
child word frequency	0.0000	0.0000	1.10	0.2708
$F(4, 165) = 22.98, p < 0.0001$, Adjusted $R^2 = 0.3422$				
		Ethan		
(Intercept)	0.6592	0.0586	11.25	0.0000***
word length	-0.0587	0.0160	-3.68	0.0003***
child minimal pairs	0.0103	0.0020	5.29	0.0000***
child phonotactic probability	-6.3902	2.9075	-2.20	0.0295*
child word frequency	0.0000	0.0001	0.12	0.9015
$F(4, 150) = 35.4, p < 0.0001$, Adjusted $R^2 = 0.4719$				
		Lily		
(Intercept)	0.7275	0.0598	12.17	0.0000***
word length	-0.0304	0.0164	-1.86	0.0640
child minimal pairs	0.0096	0.0019	5.04	0.0000***
child phonotactic probability	-6.4253	3.2096	-2.00	0.0463*
child word frequency	-0.0000	0.0000	-1.46	0.1442
$F(4, 281) = 24.6, p < 0.0001$, Adjusted $R^2 = 0.2488$				
		Naima		
(Intercept)	0.7144	0.0448	15.94	0.0000***
word length	-0.0365	0.0108	-3.38	0.0008***
child minimal pairs	0.0066	0.0015	4.34	0.0000***
child phonotactic probability	-5.9205	1.9780	-2.99	0.0029**
child word frequency	-0.0000	0.0000	-0.46	0.6461
$F(4, 381) = 29.12, p < 0.0001$, Adjusted $R^2 = 0.2261$				
		William		
(Intercept)	0.7435	0.0549	13.55	0.0000***
word length	-0.0497	0.0141	-3.53	0.0005***
child minimal pairs	0.0110	0.0018	6.28	0.0000***
child phonotactic probability	-7.8649	2.3223	-3.39	0.0009***
child word frequency	-0.0000	0.0001	-0.38	0.7022
$F(4, 157) = 36.24, p < 0.0001$, Adjusted $R^2 = 0.4669$				
		Violet		
(Intercept)	0.7206	0.0894	8.06	0.0000****
word length	-0.0603	0.0279	-2.16	0.0330*
child minimal pairs	0.0102	0.0025	4.05	0.0001***
child phonotactic probability	-3.1716	3.8438	-0.83	0.4110
child word frequency	0.0001	0.0001	0.87	0.3883
$F(4, 113) = 18.26, p < 0.0001$, Adjusted $R^2 = 0.3711$				

Table 4.5: Linear regression results for the six children for word production accuracy.

Statistical modeling. For all children except Lily, the multiple regression models show that word length, the number of minimal pairs, and phonotactic probability are predictive of production accuracy, while word frequency is not. For Lily, word length is not a significant predictor, but minimal pairs and phonotactic probability are significant like the the models for the other five children. The results are summarized in Table 4.5.

4.4.2 Phoneme level production

Production accuracy was also quantified on the phoneme level for each child in the Providence Corpus. In this section, I look at minimal pair counts and phoneme frequency as predictors as phoneme production accuracy.

4.4.2.1 Methods

Phoneme level production accuracy was measured for each phoneme for each of the six children in the Providence Corpus. The quantification of phoneme production accuracy was more difficult than word level accuracy. When producing many words, children frequently omitted parts of the words and simplified consonant clusters. These words should not be discounted when measuring phonemic production accuracy. In cases with missing phonemes, the produced phonemes needed to be best matched with the target forms. This was done by converting the phonetic and phonemic transcription to a templatic representation in terms of sound type and syllable structure. For example, the word "pop" has the target form /pɑp/, and this would be converted to *CVC*. If the child produces [bɑ] for *pop*, this production would be converted to *CV*. The converted CV matches with the first two letters of CVC, and therefore [b] is compared to /p/, and [ɑ] is compared to /ɑ/, while the final /p/ is ignored. Individual minimal pair counts on the phoneme level and phoneme type frequency for each child are taken from the Section 4.3. The effects of minimal pairs and phoneme frequency on phoneme production accuracy are investigated through visualization and confirmed via linear regressions. Because of the high correlation between type frequency and token frequency (Pearsons's $r = 0.872$, $p < 0.0001$), only type frequency is used in these

linear regression models.

4.4.2.2 Results

Minimal pairs. Figure 4.10 visualizes the relationship between phoneme production accuracy and minimal pair counts for each child. For all six children, there is a clear trend between the number of minimal pairs a phoneme has and how accurately it is pronounced. Certain phonemes, like /r/ and /ð/, appear to be especially difficult even though they have many minimal pairs.

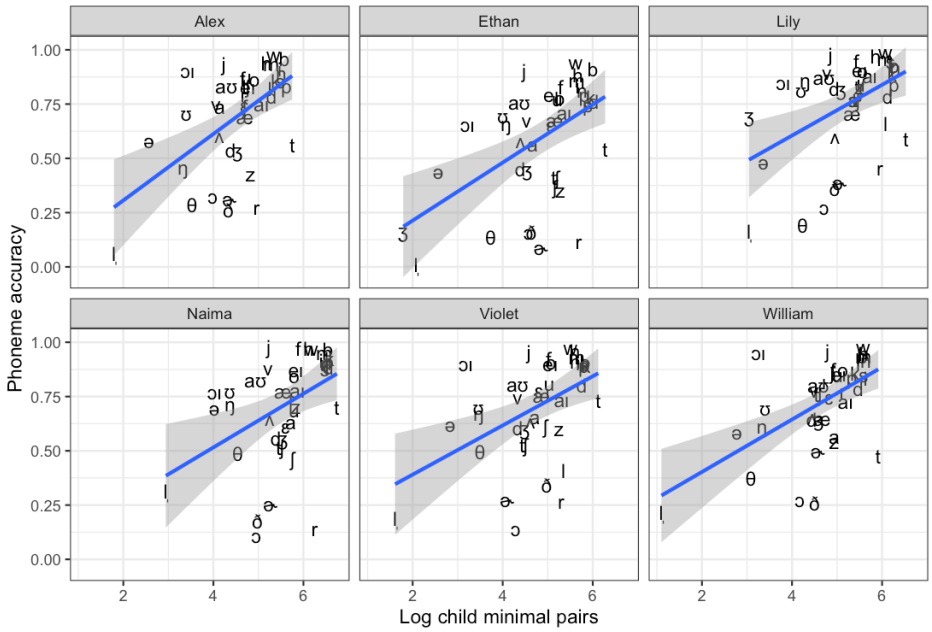
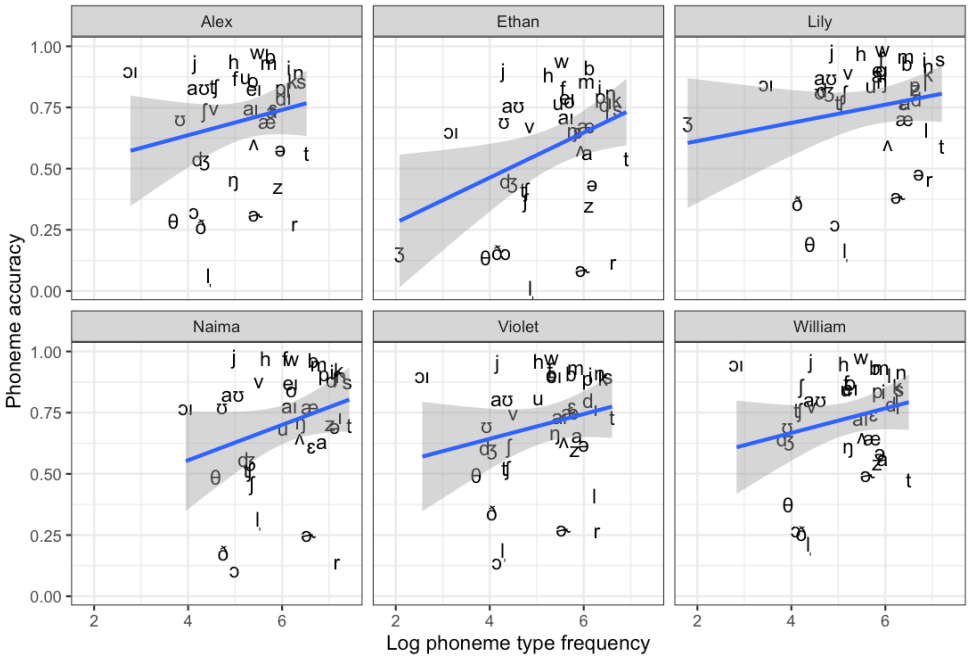


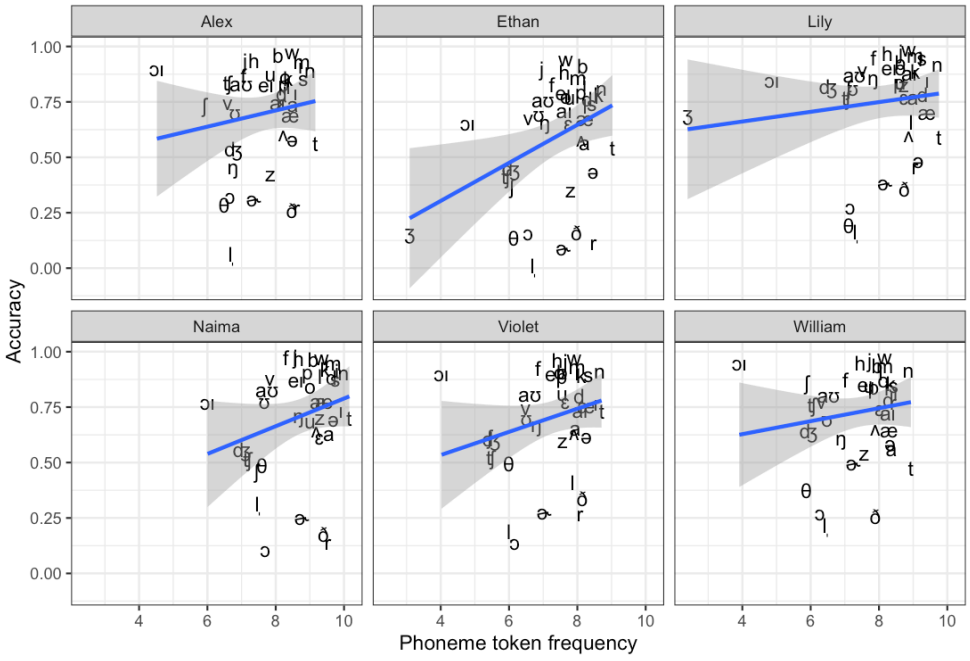
Figure 4.10: Child phoneme production accuracy and the number of minimal pairs.

Phoneme type frequency. There is a slight tendency for phoneme type frequency also. More frequent phonemes appear to be more accurately produced for some children, but the relationship is a lot weaker than minimal pairs.

Phoneme token frequency. Similar to type frequency, there is a slight tendency for phoneme token frequency. More frequent phonemes appear to be more accurately produced for some children, but the relationship is a lot weaker than minimal pairs.



(a) Child phoneme type frequency and child phoneme production accuracy.



(b) Child phoneme token frequency and child phoneme production accuracy.

Figure 4.11: Frequency and production accuracy

	Estimate	Std. Error	t value	Pr($>\|t\|$)
		Alex		
(Intercept)	0.5475	0.0636	8.61	0.0000***
child minimal pairs	0.0021	0.0006	3.62	0.0009***
child type frequency	-0.0005	0.0003	-1.75	0.0890
$F(2, 36) = 7.306, p = 0.00217$, Adjusted $R^2 = 0.2492$				
		Ethan		
(Intercept)	0.4375	0.0698	6.27	0.0000***
child minimal pairs	0.0013	0.0005	2.70	0.0103*
child type frequency	-0.0002	0.0002	-0.97	0.3368
$F(2, 37) = 5.214, p = 0.01012$, Adjusted $R^2 = 0.177$				
		Lily		
(Intercept)	0.6410	0.0602	10.65	0.0000***
child minimal pairs	0.0008	0.0003	2.93	0.0058**
child type frequency	-0.0002	0.0001	-1.61	0.1164
$F(2, 37) = 4.692, p = 0.01528$, Adjusted $R^2 = 0.1592$				
		Naima		
(Intercept)	0.5554	0.0678	8.19	0.0000***
child minimal pairs	0.0006	0.0002	2.78	0.0086**
child type frequency	-0.0001	0.0001	-0.88	0.3824
$F(2, 36) = 5.116, p = 0.01108$, Adjusted $R^2 = 0.178$				
		Violet		
(Intercept)	0.5729	0.0611	9.38	0.0000***
child minimal pairs	0.0014	0.0004	3.15	0.0033**
child type frequency	-0.0003	0.0002	-1.37	0.1778
$F(2, 36) = 5.858, p = 0.006272$, Adjusted $R^2 = 0.2036$				
		William		
(Intercept)	0.5835	0.0576	10.14	0.0000***
child minimal pairs	0.0015	0.0004	3.37	0.0018**
child type frequency	-0.0003	0.0002	-1.26	0.2161
$F(2, 36) = 6.679, p = 0.003411$, Adjusted $R^2 = 0.2301$				

Table 4.6: Linear regression results for the six children for phoneme production accuracy. The difference in degree of freedom is the result of the phoneme /ʒ/, which is missing in some children's production.

Statistical modeling. In order to test whether these trends are significant, linear regressions were run for each of the children, and the results are summarized in Table 4.6. For all six children, the regression results show that minimal pair counts significantly predict phoneme production accuracy for all six children, while phoneme type frequency does not. To ensure that the lack of frequency effects is not the result of collinearity between minimal pairs and frequency counts, simple linear regressions are estimated for both type frequency (accuracy∼type frequency) and token frequency (accuracy∼token frequency) for each child.

Neither token or type frequency is a significant predictor in all twelve simple regression models.

4.5 Discussion

This chapter has two main goals. The first is to quantify the amount of minimal pair cues in the interaction between parents and young children, and the second goal is to investigate the lexical and sub-lexical factors in phonological acquisition. The results from both parts have important implications for the study of phonological acquisition.

4.5.1 Minimal pair cues in parental input and child speech

To accomplish the first goal, a detailed examination of minimal pairs in parental and child speech was carried out. To address the potential concern that very young children might not be able to make use of functional words or morphologically complex words, minimal pair counts were collected with six levels of word category exclusion. The results show that when certain word categories are excluded, the relative numbers of minimal pairs for each phoneme remain similar. There is very high correlations between minimal pair counts with varying degrees of word category exclusion as well as between parent and child counts. These results show that although the exclusion of certain word categories might make sense from the point of view of the child's linguistic ability, word exclusion in minimal pair counting should not make a huge impact on the results of any statistical analysis due to the high correlations between the different counts.

Moreover, the results from the minimal pair counts show that parental speech contain a surprising amount of words that differ by one segment even within a single hour. Even though some words that form minimal pairs might be rare in child-directed speech, many minimal pairs are highly frequent. Within the first few years of life, children are constantly exposed to minimal pair cues in natural speech, and the abundance of minimally-contrastive words allows the child to refine their phonological knowledge. Additionally, a pairwise phoneme minimal pair analysis shows that contrastive words exist for most pairs of phonemes as

well, with exception of the phonotactically limited /ʒ/. There is clearly copious information for the learner to acquire phonological contrasts based on lexical contrast as quantified by minimal pairs.

4.5.2 Lexical contrast and minimal pairs

As reviewed in Section 4.1.2.4, developmental and historical studies of language use concepts related to minimal pairs commonly associated with phonological analysis. Studies in first language acquisition often look at the effects of phonological neighborhood density. The computation of phonological neighborhood density is similar to minimal pairs except that it typically includes words that differ through the addition or deletion of a phoneme. Although phonological analysis often only uses words of the same length in minimal pair analysis, the definition of minimal pairs does not exclude contrast with a "null" phonological unit. Functional load, most often used in diachronic studies, is in fact mostly measured through the number of minimal pairs a phoneme distinguishes.

While there is very little difference in quantification of phonological neighborhood density, functional load, and minimal pairs, these ideas are conceptualized and used differently in the literature. To advance our understanding of language acquisition and sound change, it is important to recognize that these different terms in fact measure the same thing: the amount of lexical contrast a phonological unit carries in the lexicon. The continued separation of these ideas is unnecessary and will only impede future efforts to better understand the interaction between lexical and phonological development and change.

4.5.3 Minimal pairs and phonological learning

To address the second goal of this chapter, word production and phoneme accuracy were used as approximate measures for phonological acquisition. On both the word level and the phoneme level, minimal pair counts are significant predictors of production accuracy as shown by linear regression results. This holds true for all six children in the corpus, based on their individual accuracy and minimal pair data. The consistency of the results suggest

that minimal pairs are in fact an important cue for phonological acquisition.

In learning native sound categories, acoustic distributions and acoustic salience can play an important role in perceptual tuning. However, once linguistic cues are available in the form of contrastive lexical items, the learner can rely on these cues to acquire phonological distinctions. A lexical contrast model of learning does not require the learner to hear many words frequently to acquire a contrast; this model only requires the learner to have enough experience to understand that two words have distinct meanings. Although the learner may acquire a contrast on the most frequent pair "go" and "know" faster and earlier, the learner can just as well learn from "ball" and "call" as soon as they acquire these words and understand that these words have distinct meanings.

Studies in other domains of language acquisition suggest that it is often not the quantity of input that matters, but rather the quality. For instance, in word segmentation, while infants can make use of statistical cues (Saffran et al., 1996), they used speech cues such as stress rather than statistical cues when both cues are available (Johnson and Jusczyk, 2001; Peña et al., 2002; Thiessen and Saffran, 2003; Yang, 2004; Shukla et al., 2011). Similarly, for word learning, while word frequency and the amount of input clearly have an effect on the vocabulary size of the learner (Hart and Risley, 2003), the clarity of referential cues can also affect learning of new words (Cartmill et al., 2013; Trueswell et al., 2016).

The results from the minimal pair study can be related to other studies that use metrics similar to minimal pairs. Because of the similarities in phonological neighborhood and minimal pair measures on the word level, it is not surprising that phonological neighborhood density and minimal pairs make similar predictions on the word level (e.g., Carlson et al., 2014). Likewise, there is parallel between minimal pair findings here and functional load in studies of sound change. This is not surprising since sound change occurs as a result of language acquisition.

4.5.4 Phonotactic probability

The effect of phonotactic probability is investigated on the word level. For all six children, there is a negative trend – words with higher phonotactic probabilities are less accurately produced. This trend is significant for five of the children. There are several possible interpretations for this result, especially as the downward trend appears to be driven by relatively few words for all six children. It is possible that the downward trend is merely an artifact of frequency. First, since phonotactic probability employs phoneme sequence frequency as part of the calculation, it is possible that some of the highly frequent sequences contain relatively more difficult sounds, like [ð], and [ɹ] in words like "there" and "where". Second, another contributing factor is that the gerund ending *-ing* [ɪŋ] is highly frequent morphological suffix that can inflate the phonotactic probability of words like "making", "taking", and "sitting". These results indicate that perhaps stem-level phonotactic probability should be used to better evaluate the overall effect of phonotactic probability on phonological acquisition. Also, rather than using the suggested algorithm by Vitevitch and Luce (2004), phonotactic probability based on type rather than log token frequencies may yield more insightful results. If these results are not artifacts of token frequency, the developmental interpretation would be that children are more likely to pay attention to unfamiliar sound sequences, resulting in more accurate learning of words. Further study is necessary to determine the interaction between frequency and phonotactic probability.

4.5.5 Frequency

On both the word level and the phoneme level, the lack of frequency effects is consistent. More frequent words and phonemes are not more accurately produced. Although there is a slight trend of frequency for phonemes, the trend is not statistically significant. These results clearly show that hearing a word or a sound more frequently does not lead to better acquisition results for the given word or sound, and they directly contradict previous findings in Edwards and Beckman (2008) and Beckman and Edwards (2010). The combined lack of frequency effects and significant minimal pair effects indicate that it is how a sound functions

in a lexical system that determines its acquisition trajectory.

4.5.6 Relation to the computational model

This dissertation investigates the role of lexical contrast in phonological acquisition. Because of the significant variation and overlap in the acoustic signal of distinct phonological categories, language learners must make use of additional information in the acquisition of phonological categories. The learning model outlined in Chapter 3 proposes that lexical contrast is an important cue in the acquisition of phonological categories. This learning mechanism is supported by the corpus study of child production accuracy in this chapter. Remarkably, minimal pair counts are predictive of production accuracy on both the word level and the phoneme level, and this pattern is very consistent for all six children in the Providence corpus. It is clear that phonological acquisition is more than the acquisition of phonetic patterns and that it is crucial that the learner is able to identify meaningful contrasts in the phonetic patterns from lexical cues.

4.6 Conclusion

In this chapter, I use developmental evidence to show that minimal pair cues are abundant in parental speech and that minimal pair cues, along with word length and phonotactic probability, are predictive of child production accuracy. Similar effects are not found for frequency. These results indicate that linguistically relevant cues, such as lexical contrast, play an important role in phonological acquisition.

Chapter 5

Regular Sound Change in Emergent Phonology

This chapter considers sound change through the lens of language acquisition. Sound change occurs on the individual level when the learner acquires a grammar that differs from the grammars that generated the linguistic input in their acquisition process. In studying the history of phonological systems, the Neogrammarian hypothesis, which states that sound change is regular and exceptionless, allows for the reconstruction of earlier stages of phonology using the comparative method. However, examples of apparent lexical exceptions to sound change led to a competing proposal that sound change occurs word by word through lexical diffusion. By implementing a model of vowel acquisition that explicitly controls for word frequency and the extent of allophonic variation in the input, this chapter investigates the interaction between word frequency and acoustic differences in acquisition outcomes. The results demonstrate that frequency generally has little effect on the phonological change.

5.1 Background

Diachronic change and synchronic variation are inexorably linked through language acquisition. A model of phonological acquisition therefore needs to account for how diachronic change arises from language acquisition. On the community level and across larger time scales, sound change and variation exhibit systematicity. It is important to ask how regular sound change results despite the wide range of individual variation.

5.1.1 Diachrony and language acquisition

In linguistic analysis, there is a divide between the diachronic and synchronic studies of language. The diachronic approach to language is concerned with how a linguistic system changes over time. In synchronic analysis, diachronic factors are ignored and rightfully so; during acquisition, the learner does not have access to diachronic development of their language. While this separation in approach is a useful one, diachronic and synchronic grammars are linked through language acquisition (e.g., Lightfoot, 1991; Yang, 2000; Lightfoot, 2006). Even though the learner does not have access to the linguistic history of their native language, the learner's grammar nevertheless reflect the amalgamation of successive generalizations over variable linguistic signal of the past generations.

5.1.2 The regularity of sound change and lexical diffusion

The Neogrammarian hypothesis, explicitly formulated in Osthoff and Brugmann (1878), states that sound change is regular. This assumption is the basis for much of the work in historical linguistics. In the Neogrammarian tradition, the unit of sound change is phonological. Change occurs to phonemes and phonological features, and it applies across the board. There has been a number of studies that challenge the Neogrammarian hypothesis and propose lexical diffusion as the mechanism by which sound change is implemented (Wang, 1969; Chen and Wang, 1975). Lexical diffusion posits that the unit of change is the word, rather than some phonological entity. Under lexical diffusion, sound change first occurs in some words and then spreads to other words. While the Neogrammarian hypothesis remains a fundamental working principle in historical linguistics, lexical diffusion provides an alternative to explaining the implementation of sound change. However, both proposals need to be evaluated through the lens of language acquisition.

5.1.3 Phonetics and sound change

Diachronic change arises from synchronic variation. Co-articulation is a major source of phonetic variation and thus provides the potential for sound change. Phonological contrasts

are often realized with multiple co-varying acoustic cues. This kind of co-variation is usually systematic, and listeners are aware of and can adjust for co-articulatory effects. Ohala (1983, 1993) developed a detailed model of the mechanisms involved in sound change from co-articulation. In this model, the listeners fail to recover the speaker's intended effects and identify the co-articulatory effects as inherent properties of a segment. In these listeners' subsequent productions, the misinterpreted cue will in turn serve to indicate phonological contrast, thus introducing change in the language. This model outlines how production and perception interacts to produce sound change and gives the listener a primary role in driving sound change. Since then, a number of studies have focused on the role of speech perception and production to better understand sound change.

While co-articulation is unavoidable and widespread in speech, most co-articulatory effects do not lead to sound change. Listeners are generally successful at recovering the intended phonological target by compensating for co-articulatory effects. Although certain motor movements and acoustic results can lead to similar co-articulation cross-linguistically, languages can still adopt different strategies in dealing with such co-articulation (e.g., Beddor et al., 2002; Sonderegger and Yu, 2010), resulting in the wide range of outcomes from co-articulation.

For instance, Harrington et al. (2008, 2012) studied cue shifting in co-articulatory fronting of a vowel. They found that younger speakers showed more fronting effects than older speakers, and they compensated less for co-articulatory effects. Crucially, production and perception of the shift in cue weighting are not aligned. This finding is in line with Ohala's proposal of listener-driven sound change. The same co-articulatory effects can lead to sound change in one language but stable variation in another. Tonogensis is a well-studied case of cue-shifting. Due to physiological factors, post-stop *f0* is correlated with VOT, with long VOT leading to higher *f0*. Historically, the development of tones have been attributed to the loss of VOT contrast (e.g., Hombert et al., 1979; Kim, 2004; Kang, 2014). Nevertheless, many languages that exhibit this co-articulatory difference do not become tonal. It is necessary, then, to explain the process by which a secondary cue becomes primary.

5.1.4 Research questions

This chapter uses a computational model of vowel acquisition to provide insight into the mechanisms of sound change. Specifically, I aim to answer the following questions: What are the conditions for regular sound change? How do we account for apparent lexical exceptions if we believe sound change to be phonetically driven and regular?

5.2 Methods

This chapter adapts the learning mechanism described in Chapter 3 and implement a lexical contrast based vowel acquisition model. The input to the model is explicitly controlled for the frequency of allophonic words and the degree of phonetic variation for the allophone to see how they affect the learning outcome.

5.2.1 Vowel learning model

Like the model in Chapter 3, there are two parts to the vowel acquisition model: the lexicon, which stores information about words, and phonology, which represents the learner's knowledge about vowel categories.

5.2.1.1 Lexical learning

Lexical learning occurs in a very similar fashion as described in Section 3.2.1, with some minor differences. Like in Section 3.2.1, the learner keeps track of three pieces of information for each referent: its average acoustic signal, phonological representation, and frequency (Figure 5.1). In the vowel learning model, only F1 and F2 are remembered in the acoustic representation. Also, to simplify the problem, the model assumes that all consonants have been perfectly acquired, and the consonants are represented discretely in the phonological representation. The learned vowel categories are represented as indices. For example, if five vowel categories are learned, vowels would be represented using the numbers 0, 1, 2, 3, and 4. The acquisition of a word is modeled probabilistically according to the mechanism

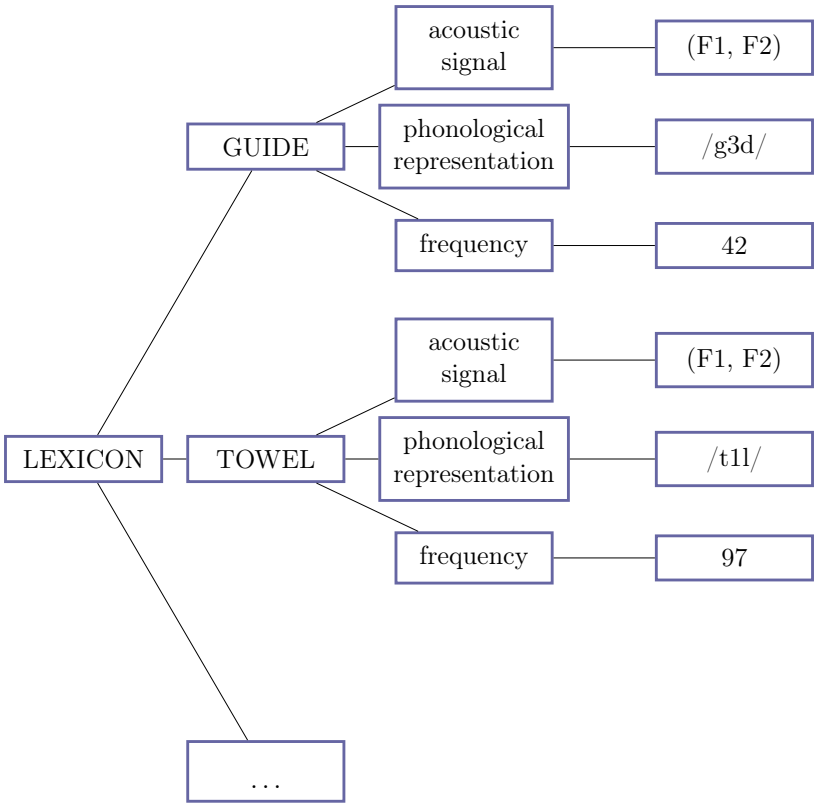


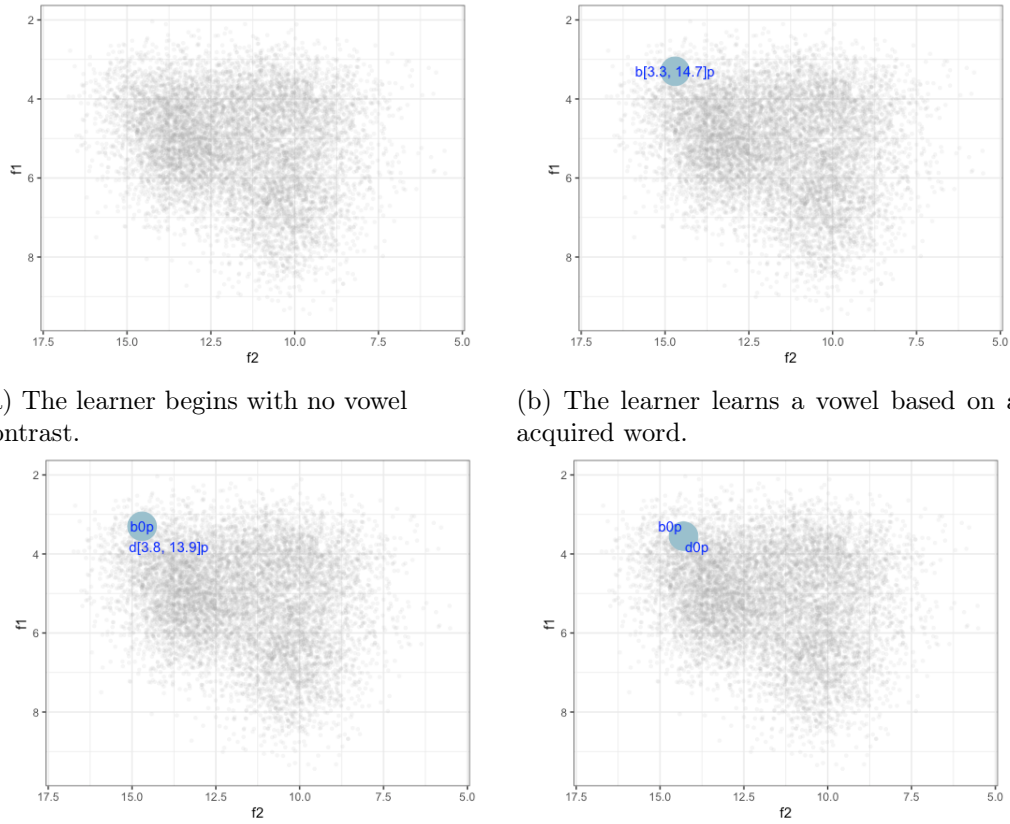
Figure 5.1: The adapted structure of the lexicon for the vowel learning model.

described in Section 3.2.1.

5.2.1.2 Phonological learning

Phonological learning is also fairly similar to the model in Chapter 3. The major difference here is that each vowel is represented as a cluster with its center calculated as the mean F1 and F2 of words assigned to this category.

Figure 5.2 illustrates the acquisition of the first vowel category. The learning begins with no vowel contrast (Figure 5.2a). Upon acquiring the first word [b(3.3, 14.7)p], the learner creates a vowel category based on the formant values of the vowel in this word (Figure 5.2b). The learner can now represent this word with a discrete vowel category and update the phonological representation of the word to /b0p/. Then, the learner acquires a second word [d(3.8, 13.9)p]. Because the onset of [d(3.8, 13.9)p] is different from /b0p/,



(a) The learner begins with no vowel contrast.

(b) The learner learns a vowel based on an acquired word.

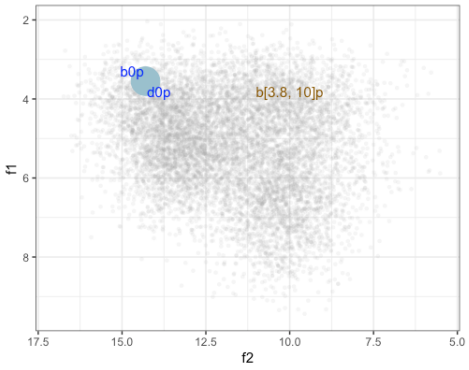
(c) The learner learns a second word.

(d) The vowel center is updated.

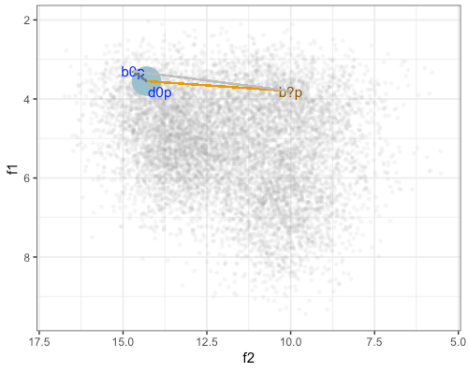
Figure 5.2: An illustration of vowel acquisition.

the lexical contrast is already represented. The learner therefore does not need to evaluate their knowledge of vowels and only needs to assign this word to the existing categories vowel category and update the category center.

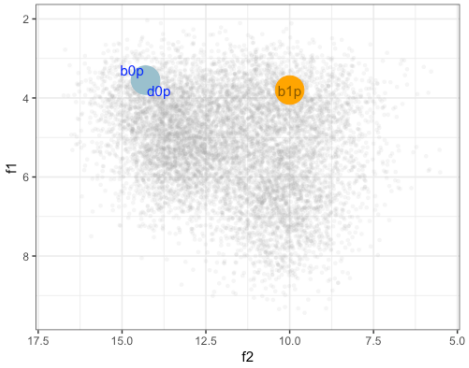
When two words with distinct referents have the same exact onset and coda, the learner needs to consider creating another vowel category (Figure 5.3a). The learner has two choices: 1) represent these two words as homophones, and 2) create distinct vowel categories for these two words. The choice is based on the relative frequency of the two words and their acoustic similarity. The processing cost of homophonic representation and contrastive representation are calculated as follows:



(a) The learner acquires a third word.



(b) The learner considers whether this is a homophone with /b0p/.



(c) The learner creates a new vowel category

Figure 5.3: A illustration of the acquisition of a second vowel.

$$C_{homophone} = \frac{\min(\text{freq}(\text{b}\textit{0}\text{p}),\ \text{freq}(\text{bVp}))}{\text{freq}(\text{b}\textit{0}\text{p})+\text{freq}(\text{bVp})} \tag{5.1}$$

$$C_{contrastive} = \frac{\max(\text{freq}(\text{b}\textit{0}\text{p}),\ \text{freq}(\text{bVp}))}{\text{freq}(\text{b}\textit{0}\text{p})+\text{freq}(\text{bVp})} \times \text{confusability} \tag{5.2}$$

where:

$$\textit{0} = \text{The learned vowel category indexed as “0”}$$

$$V = \text{A vowel that has not been assigned to a category}$$

$$\text{confusability} = \frac{d(\text{bVp},\ \textit{0})}{d(\text{b}\textit{0}\text{p},\ \textit{0}) + d(\text{b}\textit{0}\text{p},\ \text{bVp})}$$

and

$$d(a, b) = \sqrt{(F1_a - F1_b)^2 + (F2_a - F2_b)^2}$$

If $C_{\text{homophone}}$ is greater than having distinct representations $C_{\text{contrastive}}$, the learner creates another vowel category (Figure 5.3c). This process parallels the contrast determination mechanism described in Section 3.2.2.4.

5.2.2 Input generation

The input to the model is generated so that word frequencies and allophonic variations can be precisely controlled. For each learning trial, CVC words are generated by combining:

- Onset: /b p d t g k n ∅/, where ∅ indicates null onset
- Vowels: /i e a o u/, where /e/ is the vowel with an allophone
- Coda: /b p d t g k n ∅/, where ‘n$’ is a phoneme triggers an allophonic rule, and ∅ indicates null coda

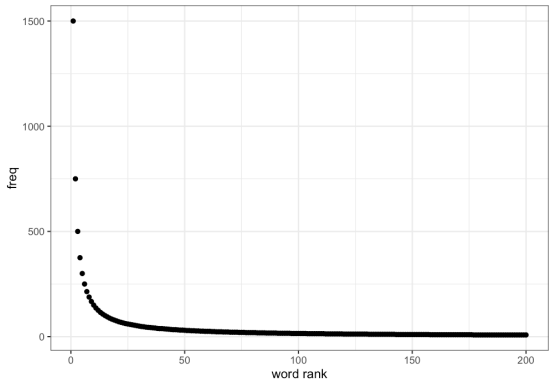
Some example words are /gan/, /nup/, /kub/, /ku∅/, /ki∅/, /nak/, /pap/, /pog/, /ted/, and /ken/. Out of the 320 possible words formed by combining the possible onsets, nuclei, and codas, 200 are randomly selected for each learning trial. The vowels are replaced by formant values (e.g., [g(F1, F2)n], [Q(F1, F2)p], [k(F1, F2)∅], [k(F1, F2)∅]).

When the vowels are replaced by formant values, the following allophonic rule is applied:

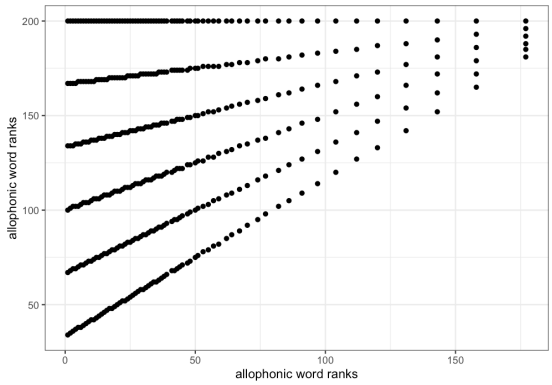
$$/\text{e}/ \rightarrow [\underline{\text{e}}] \ /\ \_\_\text{n\$}$$

That is, the vowel /e/ is realized as an allophone [e̲] when it occurs before an /n/ at the end of a word. The specific word frequency and acoustic realizations words with [e̲] are described below.

5.2.2.1 Frequency manipulations for words containing the allophone



(a) Frequencies were assigned to the ranks of the words according to the Zipfian distribution.



(b) Assigned ranks of words containing the allophonic vowel.

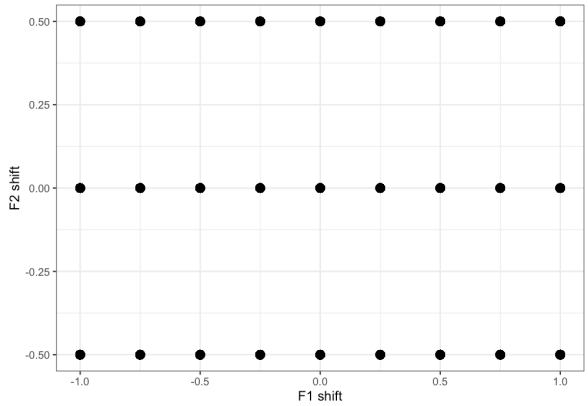
Figure 5.4: Frequency manipulations.

There are a total of 69 frequency conditions, where the words containing the allophonic [e̲] are inserted into different frequency ranks of the vocabulary. First, a maximum rank is assigned to a word with allophonic [e̲]. Next, the other words containing [e̲] are evenly distributed among the rest of the frequency ranks. For example, for frequency condition 1, a word with [e̲] is assigned the frequency rank 1, and other words containing [e̲] are spaced out evenly at ranks 51, 101, 150, and 200. For frequency condition 10, a word with [e̲] is

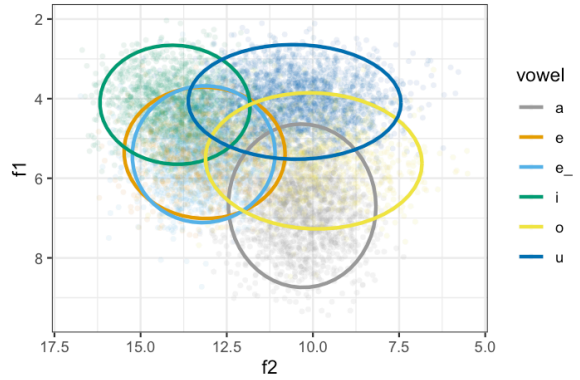
assigned frequency rank 10, and again, the rest of the words containing [e̲] are equally spaced out among the rest of the frequency ranks, at 58, 105, 152, and 200.

There are 69 total frequency conditions rather than 200 because Zipfian word frequencies are assumed (Figure 5.4a). With Zipfian distribution, words at higher ranks can share the same frequency across intervals, since frequency is a discrete number. The resulting frequency conditions is illustrated in Figure 5.4b.

5.2.2.2 Formant manipulations for words containing the allophone



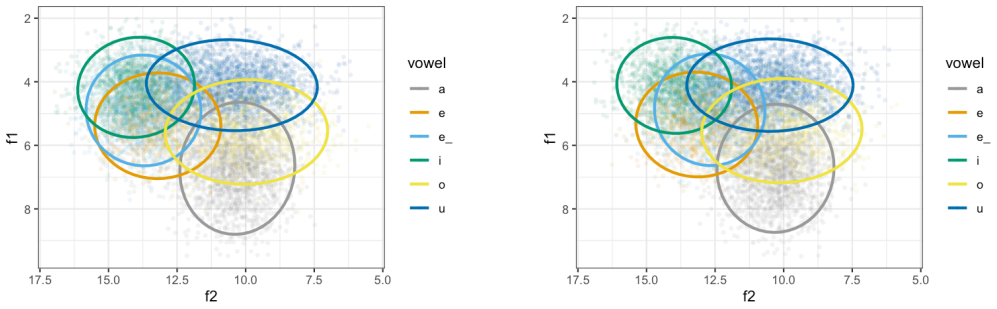
(a) F1 and F2 deviations for the allophone from the category mean for /e/.



(b) Generated input with no shift in F1 or F2.

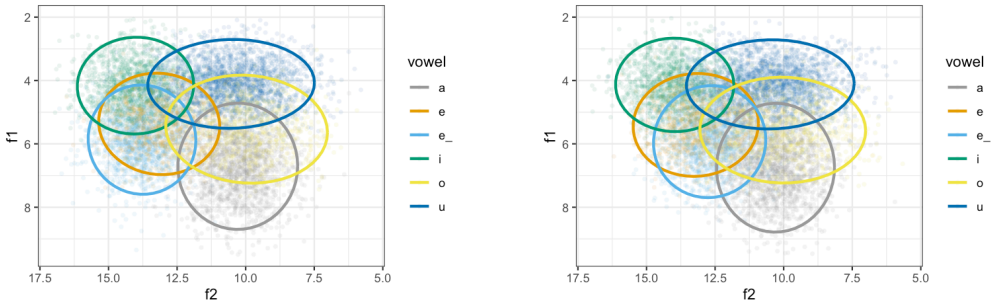
Figure 5.5: Acoustic manipulations.

As mentioned above, the allophonic rule /e/ → [e̲] / ___n$ is applied when formant values are generated for each vowel. In total, there are 27 (9 F1 × 3 F2) allophonic conditions, where [e̲] differs from [e] by some amount of shift in F1 and F2 as depicted in Figure 5.5a.



(a) Example of an allophone that is higher and more front.

(b) Example of an allophone that is higher and more back.



(c) Example of an allophone that is lower and more front.

(d) Example of an allophone that is lower and more back.

Figure 5.6: Examples of generated input with shifted in F1 or F2.

Each F1 step differs by 0.25 bark, and each F2 step differs by 0.5 bark.

The mean and standard deviations used in generating the formant values are obtained from the Philadelphia Neighborhood Corpus (Table 5.1). Figure 5.5b shows an example of generated formant values with no shift in the allophone. Figure 5.6 shows instances where the allophone has been shifted in different directions. In all cases, there is significant overlap between the formant values of the vowels.

vowel	F1 mean (bark)	F1 sd	F2 mean (bark)	F2 sd
i	4.156	0.716	13.996	1.011
e	5.390	0.770	13.187	1.057
ɑ	6.732	.951	10.291	1.009
o	5.561	0.778	10.004	1.400
u	4.103	0.669	10.513	1.414

Table 5.1: PNC formant values used in input data generation.

5.2.3 Learning trials

Each learning trial terminates after 50,000 iterations, and the learned number of vowel cluster centers and lexical representations are recorded along with the generated input for each trial. In total, there were 83,835 total trials (69 frequency steps $\times$ 27 formant steps $\times$ 45 trials).

5.3 Results

This section presents the results of the computational experiment in the following ways. First, the outcome from a typical learning trial is presented to illustrate the learned vowel representations. Next, the effects of frequency and acoustic variation on the learning outcome are visualized and analyzed.

5.3.1 An example of learning outcome

The model learns the vowel categories and assigns the learned symbolic categories in lexical representation. The learning outcome from a specific trial is presented here. The trial has an F1 shift of -0.75, F2 shift of 0.5, and allophonic frequency rank of 7. Table 5.2 shows an example of the learned lexical representations of select words in the input lexicon. As can be seen, words with the same "actual vowel" in the input are assigned to the same learned vowel category by the model. For instance, all the instances of the vowel /i/ are learned as category /4/, all the cases of /e/ are learned as /0/, and so on. In this particular trial, the

word	actual vowel	learned vowel	word freq
ni∅	i	4	574
tip	i	4	94
gig	i	4	61
pib	i	4	47
ben	e	0	8404
den	e	0	669
peb	e	0	55
keb	e	0	57
∅e∅	e	0	54
tet	e	0	178
nek	e	0	53
nɑ∅	ɑ	2	945
∅ɑk	ɑ	2	145
gap	ɑ	2	57
bɑb	ɑ	2	111
pod	o	3	73
dob	o	3	54
∅on	o	3	136
∅od	o	3	37
no∅	o	3	121
∅ug	u	1	1218
nut	u	1	260
bub	u	1	136
pug	u	1	132

Table 5.2: Example of learned lexical representations.

allophonic vowels in the words /ben/ and /den/ are assigned the same learned representation as the rest of words with /e/ despite the phonetic variation.

Where each word falls in the acoustic space is displayed in Figure 5.7. Each colored cluster represents a learned vowel category and words that belong to each category. The words are displayed using their underlying input representations rather than the learned representations. The allophonic /ben/ and /den/ are clearly removed from the rest of the /e/ cluster words, but they nevertheless are classified into the same category as other words with /e/ in this learning trial.

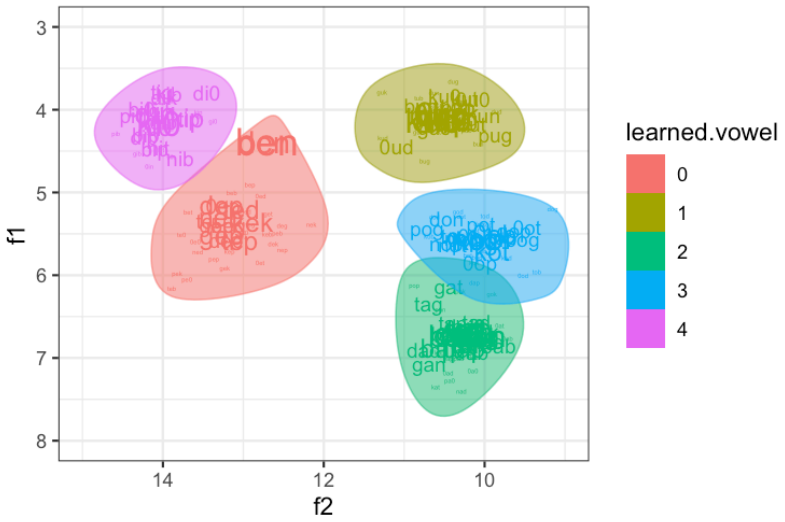


Figure 5.7: Learned word phonetics and representations. Highlighted: words containing the allophonic [e].

5.3.2 Overall learning outcome

# vowels learned	# trials	% trials
5	60638	72.33%
6	14013	16.71%
7	4611	5.50%
4	2269	2.71%
8	1486	1.77%
9	461	0.55%
10	186	0.22%
11	87	0.10%
12	37	0.04%
13	19	0.02%
3	12	0.01%
14	6	0.01%

Table 5.3: Number of vowel learned for all the learning trials.

Table 5.3 shows the number of vowel categories learned across all the frequency and acoustic conditions. The table is sorted by the percentage of trials that ended in each number of vowels, from greatest to smallest. Five vowel categories were learned for 60,638 out of 83,835 total trials, and this makes up 72.33% of all trials. Many trials also ended with 6 categories learned. For a small number of trials, the model overgenerates the number of

vowel categories and learns over 10 vowel categories. Given that this learning is unsupervised, overall the model does fairly well with discovering the right number of vowel categories.

5.3.3 Five-vowel outcomes

Out of the 60,638 learning trials where the model learned five vowels, 99.37% of the trials (60,259 trials, 71.88% of the total trials) the categories map directly onto the vowel categories (/i e ɑ o u/) in the input. This section looks more closely at the acoustic and frequency manipulations on the learning outcomes in trials that ended with five vowel categories. Both phonetic effects and phonological effects are visualized and analyzed.

5.3.3.1 Phonetic effects

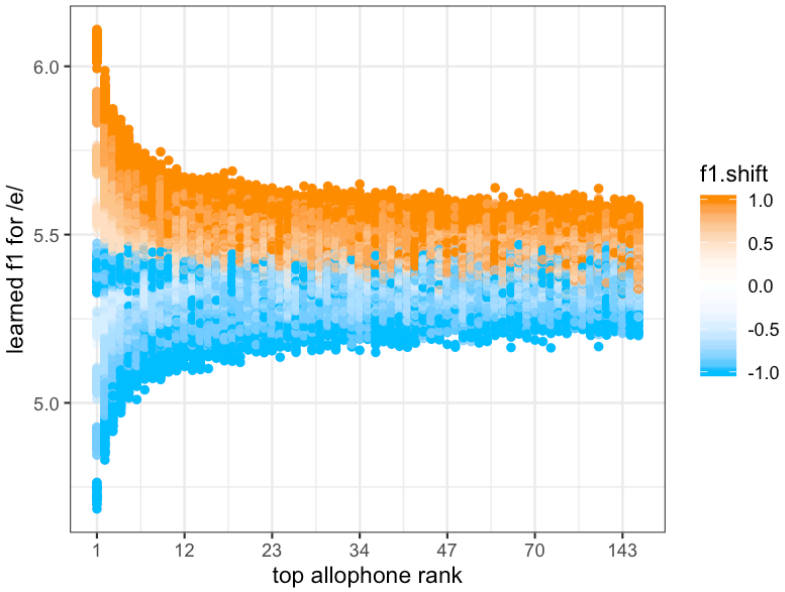


Figure 5.8: Learned F1 center for /e/ for 60,638 /i e ɑ o u/ trials.

The overall phonetic effect of the allophone frequency and acoustic variation can be observed in Figure 5.8, which shows the learned F1 center for the vowel /e/. Each dot on the plot represents the final learned F1 in each of the 60,638 trials. The input mean for the non-allophonic /e/ is 5.390 bark. The degree of phonetic shift in the allophonic F1 is represented by color, with orange indicating an increase in F1 from /e/ (lower in height) and

blue indicating a decrease in F1 Input /e/ (higher in height). Across all frequency ranks, the greater the allophonic shift, the greater the learned F1 center deviates from the input mean. Also, when the word containing the allophone is more frequent, the general shift is greater. Both the degree of phonetic variation and frequency show an effect on the phonetic learning outcome of the model.

5.3.3.2 Regular sound change

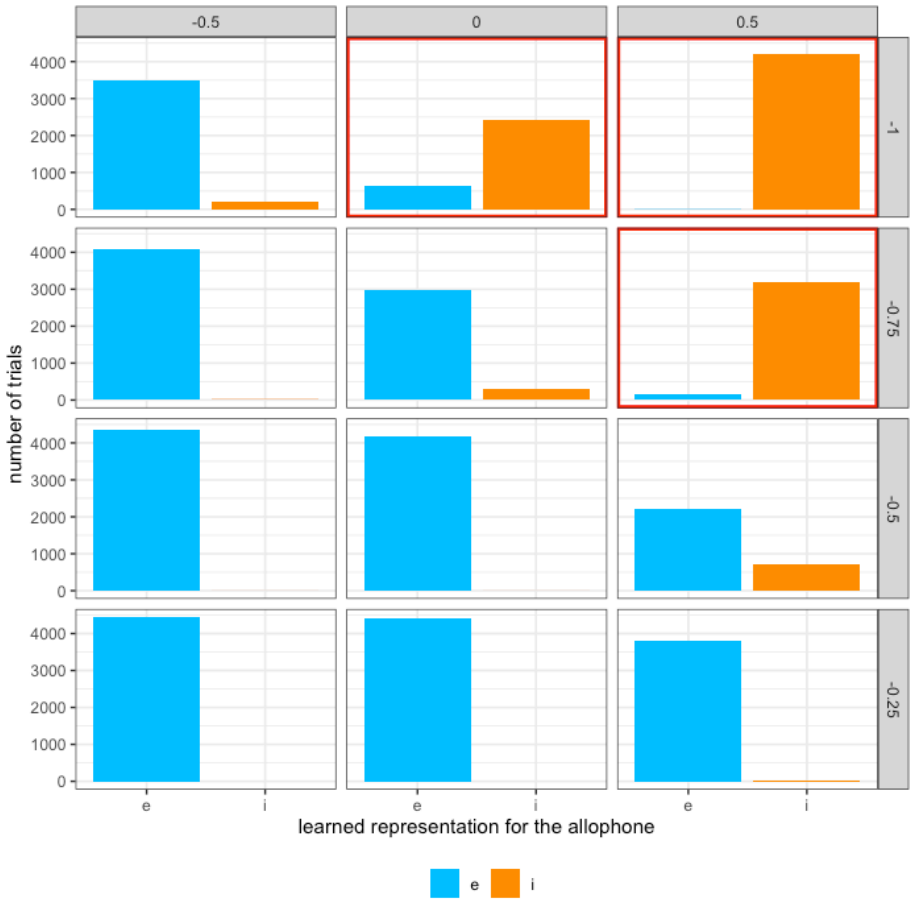


Figure 5.9: Trials in which the allophonic [e̲] is learned either as /e/ or /i/.

56,288 of the 60,259 /i e ɑ o u/ trials (93.41%) resulted in the allophonic vowel learned as exclusively /e/ or /i/. In the cases where [e̲] is learned as /e/, the allophonic rule is maintained. When [e̲] is learned as /i/, there is a phonologically conditioned phonemic change – that is, regular change has occurred in these grammars.

There is a clear observable effect of acoustic shift on the learning outcome. In Figure 5.9, each row corresponds to an F1 shift condition, and each column represents an F2 shift condition. F1 shifts of 0 and greater are omitted because the model almost learns /e/ exclusively for these conditions. The outlined are acoustic conditions where the allophone is learned as /i/ for a greater number of trials. When [e̲] is extremely high (-1 F1) and extremely front (+0.5 F2), the model almost always learns [e̲] as /i/. The model also learns [e̲] as mostly /i/ for (-0.75 F1, +0.5 F2) and (-1 F1, 0 F2) conditions.

5.3.3.3 Lexical exceptions

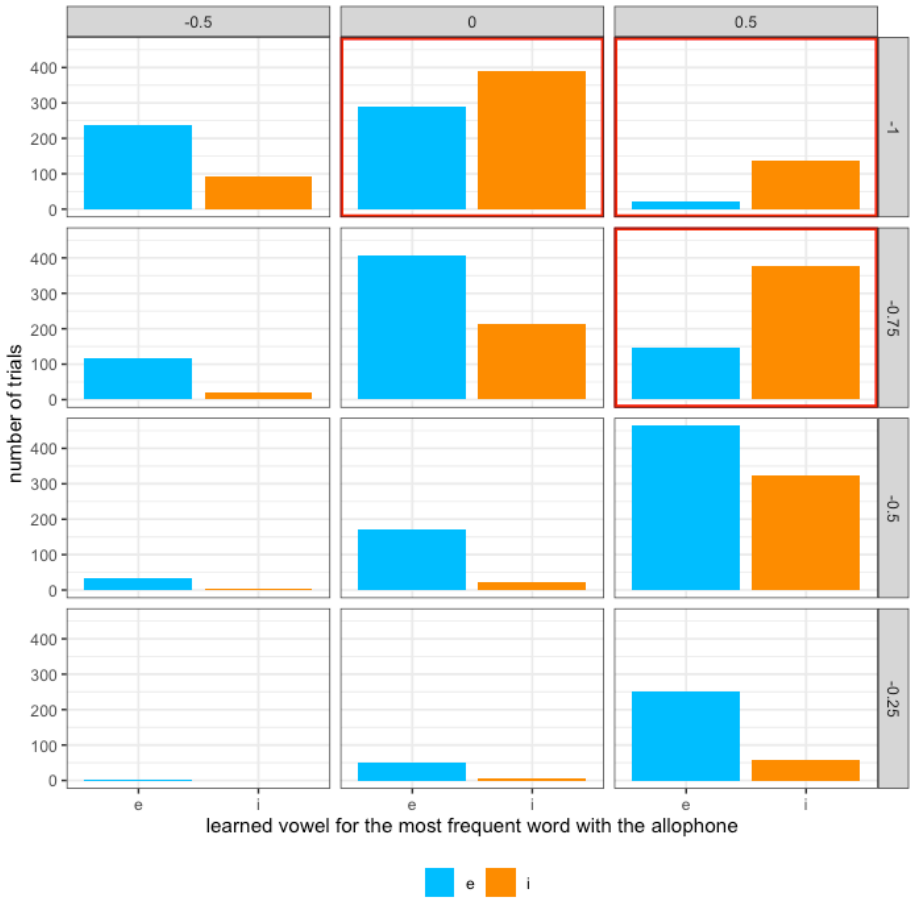


Figure 5.10: Trials in which different words with the allophonic [e̲] is assigned different representations /e/ and /i/.

There are 3971 out of 60,259 (6.58%) trials where words containing [e̲] are assigned

different representations. In each of these trials, some words with [e̲] are assigned /i/, and others are learned as /e/. Like in Figure 5.9, F1 shifts of 0 and greater are omitted because the model learned /e/ overwhelmingly for these F1 conditions. The pattern in Figure 5.10 is very similar to Figure 5.9. For acoustic conditions (-1, 0.5), (-1, 0), and (-0.75, 0), the more frequent allophonic word is assigned the representation /i/ rather than /e/. Because [e̲] in the same allophonic condition has been learned as distinct phonemes, these appear to be cases where the model has learned lexical exceptions. However, these cases are relatively rare compared to the overwhelming majority of trials that exhibited regular behavior.

5.3.3.4 Frequency

The effect of word frequency on the learned representation of [e̲] is visualized in Figure 5.11. The data is divided by F1 shift along the rows and F2 shift along the columns. Each panel shows the number of trials where the model learned [e̲] as /e/ or /i/. For most combinations of F1 and F2 combinations, there is very little pattern of word frequency. Some panels show some slight upward trend of the less common category at lower word frequencies. The only obvious trend is the panel of (-1 F1, -0.5 F2), the representations appear to be more evenly split between /e/ and /i/ for high frequency trials.

5.3.3.5 Statistical modeling

	Estimate	Std. Error	t value	Pr($>\|t\|$)
(Intercept)	0.8819	0.0080	109.74	0.0000***
frequency	-0.0003	0.0002	-1.56	0.1197
F1 shift	0.2481	0.0062	40.20	0.0000***
F2 shift	-0.2160	0.0097	-22.18	0.0000***
				Adjusted $R^2 = 0.5304$
				F(3, 1851) = 699, p $<$ 0.0000***

Table 5.4: Linear regression results for the learned representation of [e̲].

The above sections presented the results of the learned representation of [e̲]. Overall, acoustic shift appear to play an important role in whether [e̲] is acquired as /i/ or /e/

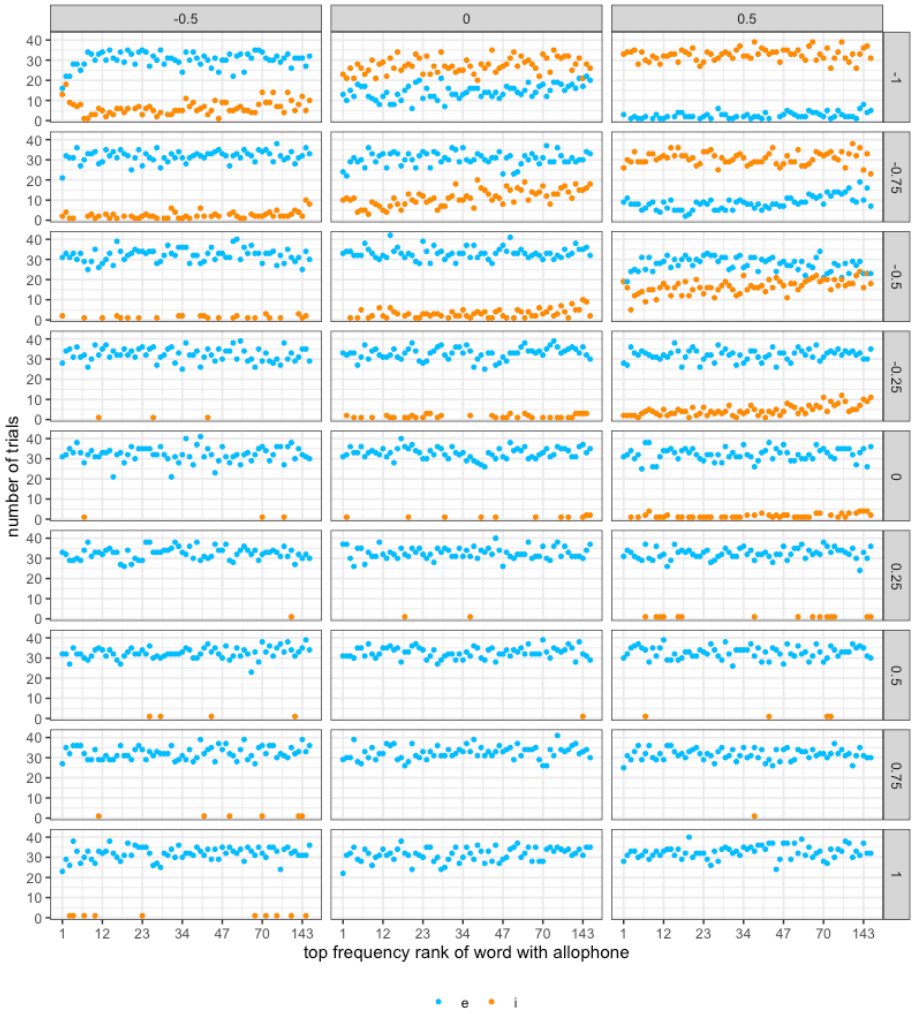


Figure 5.11: Frequency effects on the learning outcome of [e̲].

while frequency shows little to no observable trend. A multiple regression is performed to estimate the relationship between the learned representation and F1 shift, F2 shift, and word frequency. The results are presented in Table 5.4. These results confirm the observations from the visualizations that F1 and F2 have significant effects on the learned representation of [e̲], while frequency has no effect.

5.3.3.6 Other learning outcomes with five vowels

Although 99.37% of the trials with five learned vowels map onto /i e ɑ o u/, there is a small number of trials that learned vowel representations different from /i e ɑ o u/. Back vowels

appear to be more susceptible to allophonic or phonemic splits may be due to their higher variance when compared to front vowels. /o/ is highly overlapped with /u/ and /ɑ/ (Figure 5.5b).

learned vowels					# trials
ɑ	e	o	o̲	u	242
ɑ	ɑ̲	e	o	u	88
ɑ	e	o	u	u̲	39
ɑ	e	e̲	i	u	4
ɑ	e	i	u	u̲	2
ɑ	i	o	o̲	u	2
ɑ	ɑ̲	i	o	u	1

Table 5.5: Five-vowel outcomes that are not /i e ɑ o u/.

5.3.4 Learning outcomes with six and more vowels

learned vowels						# trials
ɑ	e	i	o	o̲	u	7798
ɑ	ɑ̲	e	i	o	u	2562
ɑ	e	i	o	u	u̲	1752
ɑ	e	e̲	i	o	u	1126
ɑ	e	i	i̲	o	u	669
ɑ	ɑ̲	e	o	o̲	u	51
ɑ	e	o	o̲	o̲̲	u	33
ɑ	e	o	o̲	u	u̲	14
ɑ	ɑ̲	ɑ̲̲	e	o	u	5
ɑ	ɑ̲	e	o	u	u̲	2
ɑ	e	o	u	u̲	u̲̲	1

Table 5.6: Summary of trials that learned six vowels.

In 16.71% of the trials, the model learned six vowel categories. The learned categories and their counts are presented in Table 5.6. Similar to the non-/i e ɑ o u/ five-vowel trials, the back vowels /ɑ o u/ are the ones most likely to split. Again, this might be due to the higher variance and degree of acoustic overlap in the back vowels.

5.4 Discussion

This chapter presents a model of vowel acquisition to investigate frequency and acoustic effects of allophonic variation on phonological acquisition. Overall, the model is generally successful at learning the appropriate number of vowels. There is a significant acoustic effect. While the frequency of the words containing the allophone affects the learned acoustics of the vowel category, it has little effect on the learned abstract representations of the words. Overall, the learning results support the Neogrammarian hypothesis that sound change is regular. The learning mechanism described in this chapter produces sound change that is phonetically gradual and lexically abrupt.

5.4.1 Vowel learning

First, the model is successful in adapting the learning mechanism in mechanism as Chapter 3 to acquisition on the segmental level. The vowel acquisition model shares many advantages and properties as the model in Chapter 3, especially in its nonparametric and unsupervised approach to phonological category acquisition. Across a large number of independent trials, the model succeeds in learning vowel categories in the input 71.88% of the time. There are very few instances where the model learned fewer than five vowel categories. Most of the error comes from creating additional categories, likely due to the high variance in some of the vowel categories. The vowel categories that are most likely to split are /o/, /ɑ/, and /u/. The model can be improved to better handle variance and overlap in these vowel categories.

5.4.2 Phonetic change

The learning results show both frequency and phonetic effects on the phonetic representation of the category /e/. When the allophone is more frequent, the overall category is pulled more in the direction of the allophone. Lexical diffusion can be interpreted as through the lens of phonetic change. Words with stronger co-articulation can appear to be ahead of other words in terms of change, but once the grammar calibrates around the new category

center, the phonetic shift will spread to the rest of the words.

5.4.3 Phonological change

By looking at the learned phonological representation of the words, it is possible to observe patterns of phonological change learned by the model. The learned representations of the allophonic vowel is overwhelmingly regular. In the majority of the cases, the words containing the allophonic [e̲] is learned categorically as /i/ or /e/. These learning results align with patterns observed in the historical development of language. When [e̲] is learned as /e/, the learner is maintaining the allophonic rule, and when [e̲] is learned as /i/, there is sound change on the individual level of the learner. The learner has assigned a different phonological representation to [e̲] from the grammar that generated the input. Since these are the majority of the outcome of the learning, this model in part explains how the regularity of sound change might arise.

In 6.58% of the cases, the model learns the allophonic [e̲] as /i/ in some words but /e/ in others. In these cases, the learner has effectively acquired lexical exceptions. However, these are very few in number when compared to most of the learning results that show regular outcomes. While it is possible that lexical exceptions can be observed on the individual level, grammars with lexical exceptions are rare on the community level. The learner may later adapt their grammar to be more in line with the grammar of the community. If the learner maintains the lexical exceptions, it is unlikely that these exceptions will take over as the community standards. Because historical records most likely reflect grammars on the communal rather than individual level, the observed changes across a large time scale will show regular patterns.

Overall, the modelling results indicate that the Neogrammarian hypothesis is a good working principle, since sound change is mostly categorical and word frequency generally have little effect on learning results on the phonological level.

5.4.4 Future directions

To validate the results from the model, it is important to test the learning mechanism on real cases of phonetic and phonological change. There are many possibilities for such work, both on ongoing sound changes and historical ones. Moreover, it is also important to have a community level model of sound change. The model in this chapter only produces learned results on the individual level. An interesting next step is to study how these individual grammars interact to produce sound change observed on the community level. In addition, another interesting direction of research is using the learned lexical and vowel outcomes as input to the model to simulate sound change through many generations.

5.5 Conclusion

In order to better understand how sound change occurs, it is important to have a predictive model of language acquisition. Since the function of phonological units is to provide contrast, it is worthwhile to investigate vowel acquisition in terms of lexical contrast. The contrast-based vowel acquisition model is successful at learning vowel categories, and the learning results from a large number of trials show that frequency only affects change on the phonetic level while phonological change is driven by the extent of acoustic variation.

Chapter 6

Conclusion

The goal of this dissertation is to develop and support an account of phonological category acquisition based on the interaction between lexical contrasts and acoustic distinctions. Such an account of phonological representation needs to both provide satisfactory explanations for documented phonological phenomena and hold predictive power with respect to language acquisition and sound change. The overarching hypothesis tested in this dissertation is that phonological categories emerge from the *systematic organization* of the high dimensional *acoustic space* to best accommodate the representation of *lexical contrast* in the learner's growing lexicon. Overall, the results of this dissertation support a learning mechanism by which discrete phonological categories emerge from the learning process as the learner creates meaningful divisions in the acoustic space to distinctly represent the increasing number of words in their lexicon.

6.1 Summary of contributions

The most important contribution of this dissertation is the proposal of a concrete learning mechanism that leads to the emergence of phonological categories. Previous theories of phonology often assume features to be part of Universal Grammar. However, the assumption of innate features faces many challenges in accounting for the wide range of phonological and phonetic phenomena encountered in natural language. This dissertation identifies lexical contrast as the linguistically relevant cue in phonological category acquisition and representation. The idea of phonological categories as contrastive units of word distinctions has a

long tradition in phonological analysis. Specifically, minimal pairs – words that contrast by one phonological unit – are used as a diagnostic for phonemes as the first step of establishing the phonological system of a language. However, lexical contrast has not received much attention in first language acquisition as an explanatory factor, and the work in this dissertation aims to fill this gap and motivate further research in this direction.

Using a computational model, Chapter 3 shows that the division of the high-dimensional acoustic space to accommodate the structure of lexical contrast is a viable mechanism for the acquisition of phonological categories. The results of the computational model indicate that innate features are not necessary for the acquisition of discrete phonological representations; phonological categories can emerge through a transparent mechanism where phonological contrasts can be learned from the input in a nonparametric and unsupervised fashion. The minimal assumptions and requirements of the acquisition model are a significant advance from previous computational studies in phonetic and phonological learning.

Chapter 4 presents a quantitative analysis of the lexical and sub-lexical factors in word and phoneme acquisition through a corpus study of child speech. There are two striking results from this corpus study. First, there is a surprisingly large number of minimal pairs in parental speech. Second, the number of minimal pairs significantly predicts child production accuracy on both the word level and the phoneme level, while word and phoneme frequencies have no effect on production accuracy. This pattern is consistent in the individual data for all six children in the corpus. Minimal pairs are high signal words that can help the learner draw finer boundaries for contrastive phonological units, leading to better learning outcomes for words and phonemes with more minimal pairs. These results indicate that linguistically relevant cues, such as lexical contrast, play an important role in phonological acquisition.

In Chapter 5, the acquisition model is adapted to explore the effects of acoustic variation and word frequency in sound change. First, this chapter shows that the learning mechanism outlined in Chapter 3 can be easily adapted to learn on the segmental level. Second, the modeling results display very little frequency effects. Under the assumption that lexical contrast drives phonological category learning, sound change is mostly regular, as is the

case in many well documented sound changes. The regularity of sound change can also be seen as an emergent outcome of the phonological acquisition process.

Overall, using computational and corpus methods, this dissertation provides substantial evidence that the acquisition and refinement of phonological units depends on the structure of the lexicon.

6.2 Future directions

The goal of linguistics to better understand the cognitive structures and mechanisms that underlie linguistic competence. This dissertation has argued for a closer examination of Universal Grammar, and the results of the computational and corpus studies indicate that some properties of languages that have long been assumed to be innate may in fact be emergent. It is important to continue to address the larger theoretical question of which properties of language are emergent and how they arise through the interaction between Universal Grammar and primary linguistic data that a learner is exposed to.

This dissertation focuses on the emergence of phonological categories and presents a view of Universal Grammar as a learning mechanism rather than a set of predefined innate structures or features. Similar questions apply to other levels of linguistic representation. The general concrete and abstractness of representation exist on the morphological, syntactic, and semantic levels despite similarly noisy signal in the linguistic input. The stability of grammar in the short run indicates language acquisition must be constrained such that learners generally reach comparable conclusions about the structure of their ambient language. However, variation and change necessitates a linguistic learning mechanism that is flexible and allows for innovations. The exact nature of Universal Grammar remains an open question, and more studies of language development are needed in order to better distinguish between the innate and emergent properties of language.

Appendix A

Additional Analyses for Chapter 4

A.1 Categorical and gradient word accuracy

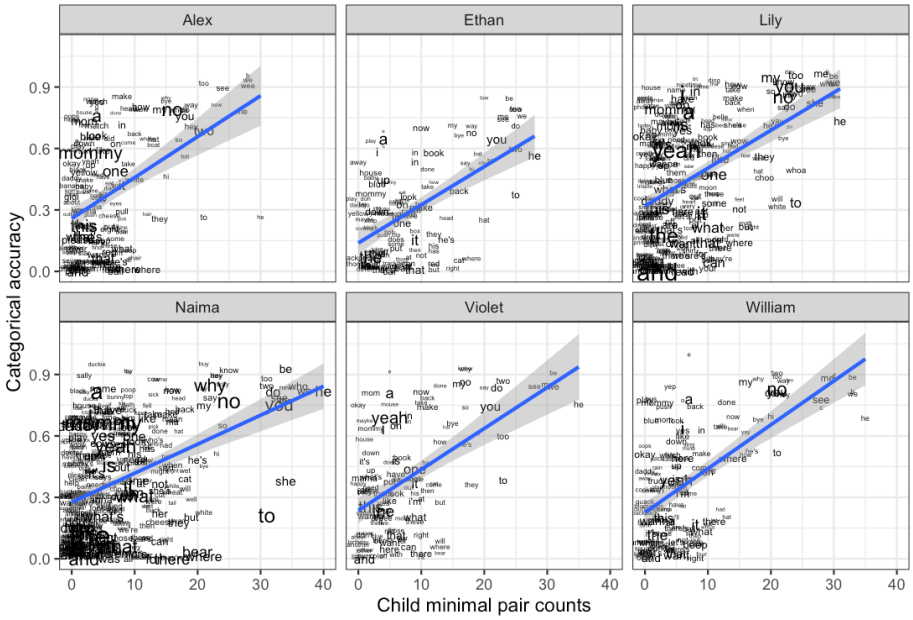
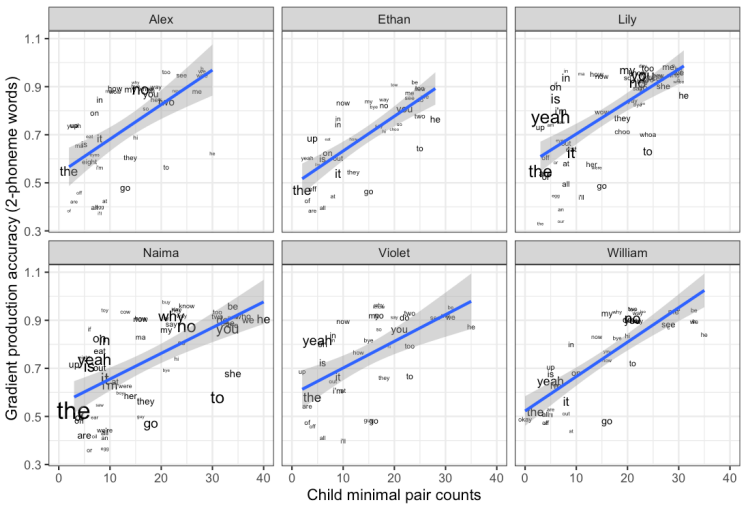
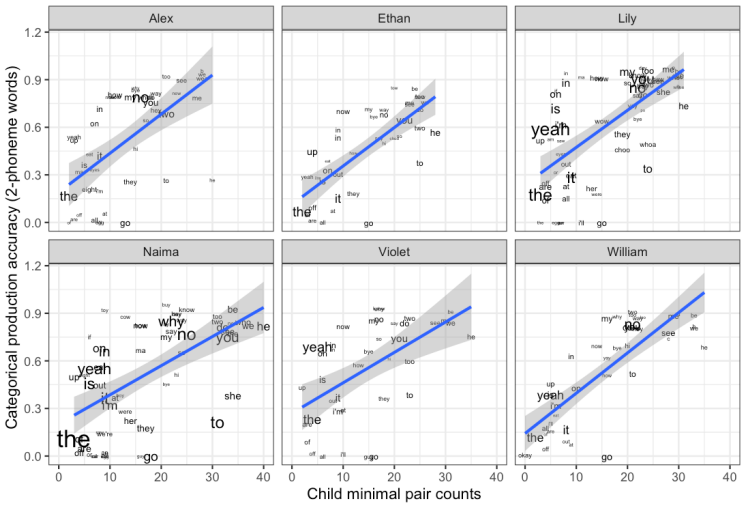


Figure A.1: Child minimal pair counts and gradient word production accuracy for the six children for all words. There is an overall trend that more minimal pairs indicate better production accuracy.

Figure A.1 visualizes the production accuracy if categorical word accuracy is used instead of gradient word accuracy. The trend is very similar to Figure 4.7a in Chapter 4. Figure A.2 shows a comparison for these two measures for 2-phoneme words.



(a) Minimal pairs and gradient accuracy for 2-phoneme words.

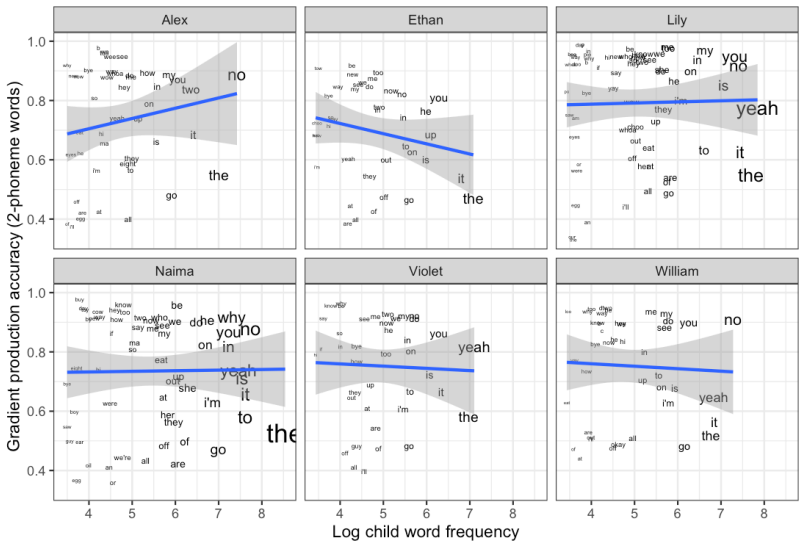


(b) Minimal pairs and categorical accuracy for 2-phoneme words.

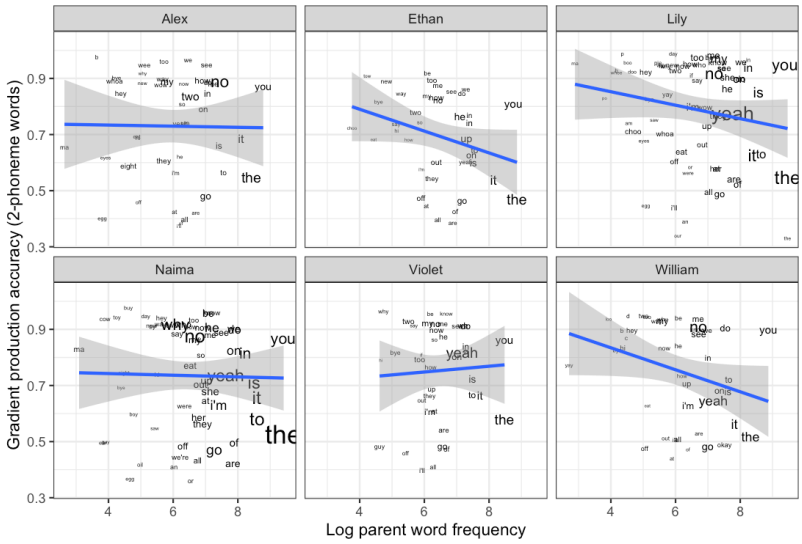
Figure A.2: Gradient vs. categorical word accuracy measures for 2-phoneme words.

A.2 Child vs. parental frequencies

Chapter 4 uses child word frequencies in visualizations and analyses. Figure A.3 shows that similar trends (or lack of trends) can be observed with either child of parental word frequencies.



(a) Child frequencies.



(b) Parent frequencies.

Figure A.3: Child word frequencies vs. parental word frequencies and production accuracy.

A.3 Statistical models with parental counts

The statistical analyses in Chapter 4 use measures derived from child production. Here I present the same models with measures from parental speech.

A.3.1 Phoneme accuracy

	Estimate	Std. Error	t value	Pr(> |t|)
		Alex		
(Intercept)	0.5691	0.0645	8.82	0.0000***
parental minimal pairs	0.0006	0.0002	3.48	0.0013**
parental type frequency	-0.0002	0.0001	-1.96	0.0572
$F(2, 36) = 6.311, p = 0.004473$, Adjusted $R^2 = 0.2184$				
		Ethan		
(Intercept)	0.4278	0.0687	6.23	0.0000***
parental minimal pairs	0.0006	0.0002	3.30	0.0022**
parental type frequency	-0.0002	0.0001	-1.41	0.1669
$F(2, 37) = 6.722, p = 0.003234$, Adjusted $R^2 = 0.2269$				
		Lily		
(Intercept)	0.6408	0.0606	10.57	0.0000***
parental minimal pairs	0.0003	0.0001	3.10	0.0036**
parental type frequency	-0.0001	0.0000	-1.89	0.0673
$F(2, 37) = 4.941, p = 0.01253$, Adjusted $R^2 = 0.1681$				
		Naima		
(Intercept)	0.5544	0.0683	8.11	0.0000***
parental minimal pairs	0.0004	0.0001	3.11	0.0037**
parental type frequency	-0.0001	0.0001	-1.26	0.2140
$F(2, 36) = 5.535, p = 0.008018$, Adjusted $R^2 = 0.1927$				
		Violet		
(Intercept)	0.5762	0.0667	8.64	0.0000***
parental minimal pairs	0.0004	0.0001	2.96	0.0055**
parental type frequency	-0.0001	0.0001	-1.62	0.1134
$F(2, 36) = 4.592, p = 0.01674$, Adjusted $R^2 = 0.159$				
		William		
(Intercept)	0.5938	0.0570	10.42	0.0000***
parental minimal pairs	0.0006	0.0002	3.39	0.0017**
parental type frequency	-0.0002	0.0001	-1.53	0.1337
$F(2, 36) = 6.325, p = 0.004424$, Adjusted $R^2 = 0.2189$				

Table A.1: Linear regression results for the six children for phoneme production accuracy using parental measures.

A.3.2 Word accuracy

	Estimate	Std. Error	t value	Pr(> |t|)
		Alex		
(Intercept)	0.7497	0.0644	11.64	0.0000***
word length	-0.0530	0.0184	-2.87	0.0046**
parental minimal pairs	0.0061	0.0014	4.45	0.0000***
parental phonotactic probability	-7.9536	3.9122	-2.03	0.0437*
parental word frequency	-0.0000	0.0000	-1.11	0.2676
$F(4, 165) = 21.03, p < 0.0001$, Adjusted $R^2 = 0.3216$				
		Ethan		
(Intercept)	0.6601	0.0527	12.52	0.0000***
word length	-0.0594	0.0150	-3.96	0.0001***
parental minimal pairs	0.0064	0.0010	6.33	0.0000***
parental phonotactic probability	-5.7943	2.9031	-2.00	0.0478*
parental word frequency	-0.0000	0.0000	-1.24	0.2185
$F(4, 150) = 40.73, p < 0.0001$, Adjusted $R^2 = 0.5079$				
		Lily		
(Intercept)	0.7448	0.0563	13.23	0.0000***
word length	-0.0367	0.0153	-2.41	0.0168*
parental minimal pairs	0.0054	0.0010	5.48	0.0000***
parental phonotactic probability	-5.8737	3.3094	-1.77	0.0770
parental word frequency	-0.0000	0.0000	-2.28	0.0236*
$F(4, 281) = 26.3.6, p < 0.0001$, Adjusted $R^2 = 0.262$				
		Naima		
(Intercept)	0.7213	0.0434	16.63	0.0000***
word length	-0.0441	0.0108	-4.09	0.0001***
parental minimal pairs	0.0048	0.0010	4.64	0.0000***
parental phonotactic probability	-2.5048	1.8948	-1.32	0.1870
parental word frequency	-0.0000	0.0000	-1.73	0.0844
$F(4, 381) = 29.17, p < 0.0001$, Adjusted $R^2 = 0.2264$				
		Violet		
(Intercept)	0.7702	0.0882	8.74	0.0000***
word length	-0.0657	0.0277	-2.37	0.0195*
parental minimal pairs	0.0059	0.0016	3.72	0.0003***
parental phonotactic probability	-6.6800	4.5500	-1.47	0.1448
parental word frequency	-0.0000	0.0000	-0.21	0.8339
$F(4, 113) = 17.17, p < 0.0001$, Adjusted $R^2 = 0.356$				
		William		
(Intercept)	0.7542	0.0536	14.08	0.0000***
word length	-0.0647	0.0145	-4.47	0.0000***
parental minimal pairs	0.0072	0.0012	5.90	0.0000***
parental phonotactic probability	-1.6196	3.0817	-0.53	0.5999
	-0.0000	0.0000	-1.86	0.0648
$F(4, 157) = 33.61, p < 0.0001$, Adjusted $R^2 = 0.4476$				

Table A.2: Linear regression results for the six children for word production accuracy.

A.4 Collinearity of predictors

Since the measures used in the linear regression models could be correlated, variance inflation factors are used to quantify any multicollinearity in the models. While there is some correlation between the independent variables, the correlations are not large.

	word length	minimal pairs	phonotactic probability	word frequency
Alex	1.75	1.54	1.21	1.11
Ethan	2.28	1.67	1.18	1.33
Lily	2.59	1.88	1.43	1.13
Naima	2.15	1.71	1.26	1.10
Violet	2.77	1.81	1.61	1.26
William	1.85	1.56	1.11	1.18

Table A.3: VIFs for models with child measures (Table 4.5).

	word length	minimal pairs	phonotactic probability	word frequency
Alex	2.05	1.57	1.31	1.12
Ethan	2.15	1.55	1.39	1.15
Lily	2.30	1.70	1.36	1.12
Naima	2.13	1.63	1.34	1.13
Violet	2.66	1.69	1.69	1.21
William	1.89	1.52	1.20	1.13

Table A.4: VIFs word accuracy models with parental measures (Table A.2).

Bibliography

Abramson, A. S. and Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, pages 25–33.

Adriaans, F. and Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, 141(5):3070–3078.

Anderson, J. R., Bothell, D., Lebiere, C., and Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, 38(4):341–380.

Bailey, T. M. and Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17(2):1265–1282.

Beckman, M. E. and Edwards, J. (2010). Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology*, 1(2):319–343.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, pages 785–821.

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4):591–627.

Bergelson, E. and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9):3253–3258.

Berko, J. and Brown, R. (1960). Psycholinguistic research methods. *Handbook of research methods in child development*, pages 517–557.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In *Developmental neurocognition: Speech and face processing in the first year of life*, pages 289–304. Springer.

Best, C. T. et al. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, 167(224):233–277.

Best, C. T., McRoberts, G. W., LaFleur, R., and Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant behavior and development*, 18(3):339–350.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by english-speaking adults and infants. *Journal of experimental psychology: human perception and performance*, 14(3):345.

Blaho, S. (2008). *The syntax of phonology: A radically substance-free approach*. Universitetet i Tromsø.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boersma, P., Cohn, A., Fougeron, C., and Huffman, M. (2012). Modeling phonological category learning. *Oxford handbooks in linguistics*.

Boersma, P., Escudero, P., Hayes, R., et al. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th international congress of phonetic sciences*, volume 1013, page 1016. Barcelona.

Boersma, P. and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry*, 32(1):45–86.

Boland, J. E. and Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45(3):391–411.

Bonin, P. and Fayol, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*, 14(3):289–313.

Borden, G., Gerber, A., and Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in korean adults learning english. *Language Learning*, 33(4):499–526.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me familiar names help launch babies into speech-stream segmentation. *Psychological science*, 16(4):298–304.

Bowers, J. S., Kazanina, N., and Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, 87:71–83.

Bradlow, A. R. (1995). A comparative acoustic study of english and spanish vowels. *The Journal of the Acoustical Society of America*, 97(3):1916–1924.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training japanese listeners to identify english /r/ and /l/: Iv. some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4):2299–2310.

Brent, M. R. and Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2):B33–B44.

Caramazza, A., Costa, A., Miozzo, M., and Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6):1430.

Carlson, M. T., Sonderegger, M., and Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of memory and language*, 75:159–180.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., and Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28):11278–11283.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child development*, pages 433–448.

Chen, M. Y. and Wang, W. S. (1975). Sound change: actuation and implementation. *Language*, pages 255–281.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper & Row, New York.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, 20(03):287–333.

Clements, G. N. and Ridouane, R. (2011). *Where do phonological features come from?: cognitive, physical and developmental bases of distinctive speech categories*, volume 6. John Benjamins Publishing.

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 13(1):153–156.

Cole, R. A., Jakimik, J., and Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, 64(1):44–56.

Cowper, E. and Hall, D. C. (2015). Reductio ad discrimen: Where features come from. *Nordlyd*, 41(2):145–164.

Cristia, A., Seidl, A., and Francis, A. (2011). Phonological features in infancy. *Where do phonological contrasts come from*, pages 303–326.

Curtin, S., Fennell, C., and Escudero, P. (2009). Weighting of vowel cues explains patterns of word–object associative learning. *Developmental Science*, 12(5):725–731.

Dautriche, I., Swingley, D., and Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, 143:77–86.

De Boer, B. and Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, 4(4):129–134.

de Boysson-Bardies, B., Hallé, P., Sagart, L., and Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of child language*, 16(1):1–17.

de Boysson-Bardies, B., Sagart, L., and Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of child language*, 11(1):1–15.

de Boysson-Bardies, B. and Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2):297–319.

de Saussure, F. (1916). *Cours de linguistique générale*. Payot. Edited by Charles Bally and Albert Sechehaye.

Demuth, K., Culbertson, J., and Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, 49(2):137–173.

Dillon, B., Dunbar, E., and Idsardi, W. (2013). A single-stage approach to learning phonological categories: Insights from inuktitut. *Cognitive science*, 37(2):344–377.

Dmitrieva, O., Llanos, F., Shultz, A. A., and Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in spanish and english. *Journal of Phonetics*, 49:77–95.

Dooling, R. J., Best, C. T., and Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave/ra–la/continua by budgerigars (melopsittacus undulatus) and zebra finches (taeniopygia guttata). *The journal of the Acoustical Society of America*, 97(3):1839–1846.

Dresher, B. E. (2004). On the acquisition of phonological representations. In *First Workshop on Psycho-computational Models of Human Language Acquisition*, page 41. Citeseer.

Dresher, B. E. (2015). The arch not the stones: Universal feature theory without universal features. *Nordlyd*, 41(2):165–181.

Dresher, B. E. (2016). Contrast in phonology, 1867–1967: History and development. *Annual Review of Linguistics*, 2:53–73.

Dresher, B. E. (2017). Contrastive feature hierarchies in phonology: variation and universality. Georgetown University Round Table.

Edwards, J. and Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language learning and development*, 4(2):122–156.

Edwards, J., Beckman, M. E., and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of speech, language, and hearing research*, 47(2):421–436.

Edwards, J., Munson, B., and Beckman, M. E. (2011). Lexicon-phonology relationships and dynamics of early language development–a commentary on stoel-gammon's 'relationships between lexical and phonological development in young children'. *Journal of child language*, 38(1):35–40.

Eilers, R. E., Wilson, W. R., and Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech, Language, and Hearing Research*, 20(4):766–780.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 16(3):513–521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [rl] distinction by young infants. *Perception & Psychophysics*, 18(5):341–347.

Eimas, P. D. and Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4(1):99–109.

Eimas, P. D., Siqueland, E. R., Juscyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968):303–306.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, 24(2):143–188.

Elsner, M., Goldwater, S., Feldman, N., and Wood, F. (2013). A joint learning model of word segmentation, lexical acquisition, and phonetic variability. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 42–54.

Ernestus, M., Baayen, H., and Schreuder, R. (2002). The recognition of reduced word forms. *Brain and language*, 81(1-3):162–173.

Escudero, P., Best, C. T., Kitamura, C., and Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Frontiers in psychology*, 5.

Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and european portuguese. *The Journal of the Acoustical Society of America*, 126(3):1379–1393.

Esling, J. H. (2012). The articulatory function of the larynx and the origins of speech. In *Annual Meeting of the Berkeley Linguistics Society*, volume 38, pages 121–149.

Feldman, N. H., Griffiths, T. L., Goldwater, S., and Morgan, J. L. (2013a). A role for the developing lexicon in phonetic category acquisition. *Psychological review*, 120(4):751.

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., and Morgan, J. L. (2013b). Word-level information influences phonetic learning in adults and infants. *Cognition*, 127(3):427–438.

Fennell, C. T. and Waxman, S. R. (2010). What paradox? referential cues allow for infant use of phonetic detail in word learning. *Child development*, 81(5):1376–1383.

Fennell, C. T. and Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and speech*, 46(2-3):245–264.

Ferguson, C. A. and Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language*, 51:419–439.

Fernald, A., Perfors, A., and Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, 42(1):98.

Fernald, A., Swingley, D., and Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child development*, 72(4):1003–1015.

Fikkert, P. (1994). *On the acquisition of prosodic structure*. [Sl: sn].

Fikkert, P. and Levelt, C. (2008). How does place fall into place? *The lexicon and emergent constraints in children's developing phonological grammar. Contrast in phonology: theory, perception, and acquisition, ed. by BE Dresher & K. Rice*, pages 231–68.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3):1730–1741.

Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in english. *The Journal of the Acoustical Society of America*, 124(2):1234–1251.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, 13(3):361–377.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1):110.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition*, 22(5):1166.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological review*, 105(2):251.

Hale, M. and Reiss, C. (2000). "substance abuse" and "dysfunctionalism": current trends in phonology. *Linguistic inquiry*, 31(1):157–169.

Hale, M. and Reiss, C. (2003). The subset principle in phonology: why the tabula can't be rasa. *Journal of Linguistics*, 39(2):219–244.

Hall, D. (2007). *The role and representation of contrast in phonological theory*. PhD thesis, University of Toronto.

Hall, K. C., Blake, A., Fry, M., Johnson, K., Lo, R., Mackie, S., and McAuliffe, M. (2017). Phonological corpustools, version 1.3. `http://phonologicalcorpustools.github.io/CorpusTools/`.

Hallé, P. A. and de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19(4):463–481.

Harrington, J., Kleber, F., and Reubold, U. (2008). Compensation for coarticulation,/u/-fronting, and sound change in standard southern british: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, 123(5):2825–2835.

Harrington, J., Kleber, F., and Reubold, U. (2012). The production and perception of coarticulation in two types of sound change in progress. *Speech planning and dynamics*, pages 39–62.

Hart, B. and Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American educator*, 27(1):4–9.

Hayes, B. (2004). Phonological acquisition in optimality theory: The early stages. *Constraints in phonological acquisition*, pages 158–203.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, 23(1):65–94.

Hazan, V. and Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of phonetics*, 28(4):377–396.

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, pages 37–58.

Horváth, K., Myers, K., Foster, R., and Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of sleep research*, 24(5):503–509.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1):105–113.

Houston, D. M. and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5):1570.

Hyman, L. M. (2011). Do tones have features. *Tones and features*, pages 50–80.

Ingram, D. (1988a). The acquisition of word-initial [v]. *Language and speech*, 31(1):77–85.

Ingram, D. (1988b). Jakobson revisited: Some evidence from the acquisition of polish. *Lingua*, 75(1):55–82.

Jakobson, R. (1941). *Aphasie, Kindersprache und allgemeine Lautgesetze.* Stockholm: Almqvist & Wiksell.

Jakobson, R. (1968). *Child language, aphasia and phonological universals*, volume 72. Walter de Gruyter GmbH & Co KG.

Jakobson, R., Fant, C. G., and Halle, M. (1951). *Preliminaries to Speech Analysis. The distinctive features and their correlates.* Acoustics Laboratory MIT.

Johnson, E. K. and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of memory and language*, 44(4):548–567.

Johnson, E. K., Seidl, A., and Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PloS one*, 9(1):e83546.

Johnson, K. (2004). Massive reduction in conversational american english. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, pages 29–54. Citeseer.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language.* MIT Press, Cambridge.

Jusczyk, P. W. and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive psychology*, 29(1):1–23.

Jusczyk, P. W. and Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, 277(5334):1984–1986.

Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33(5):630.

Jusczyk, P. W., Pisoni, D. B., Reed, M. A., Fernald, A., and Myers, M. (1983). Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals. *Science*, 222(4620):175–177.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., and Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 21(1):50–54.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in seoul korean stops: A corpus study. *Journal of Phonetics*, 45:76–90.

Kemps, R., Ernestus, M., Schreuder, R., and Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90(1-3):117–127.

Kim, M. (2004). Correlation between vot and f0 in the perception of korean stops and affricates. In *INTERSPEECH*.

Kuang, J. and Cui, A. (2018a). Perceptual equivalence between co-articulated cues during a sound change in progress. Paper presented at the LabPhon 16, Lisbon.

Kuang, J. and Cui, A. (2018b). Relative cue weighting in production and perception of an ongoing sound change in southern yi. *Journal of Phonetics*, 71:194–214.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *The Journal of the Acoustical Society of America*, 70(2):340–349.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In *Developmental neurocognition: Speech and face processing in the first year of life*, pages 259–274. Springer.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22):11850–11857.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (nlm-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):979–1000.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, 9(2):F13–F21.

Kuhl, P. K., Williams, K. A., et al. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606.

Labov, W. and Rosenfelder, I. (2011). The philadelphia neighborhood corpus of ling 560 studies, 1972-2010. *With support of NSF contract*, 921643.

Ladefoged, P. (2005). Features and parameters for different purposes. In *Paper presented at the Linguistic Society of America meeting*.

Lahiri, A. and Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3):245–294.

Lake, B. M., Vallabha, G. K., and McClelland, J. L. (2009). Modeling unsupervised perceptual category learning. *IEEE Transactions on Autonomous Mental Development*, 1(1):35–43.

Law, F. and Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development*, 11(4):331–355.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, 54(5):358.

Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. (1961). The discrimination of

relative onset-time of the components of certain speech and nonspeech patterns. *Journal of experimental psychology*, 61(5):379.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

Liddell, S. K. and Johnson, R. E. (1989). American sign language: The phonological base. *Sign language studies*, 64(1):195–277.

Lightfoot, D. (1991). *How to set parameters: Arguments from language change*. Mit Press.

Lightfoot, D. (2006). *How new languages emerge*. Cambridge University Press.

Lin, Y. (2005). *Learning features and segments from waveforms: A statistical model of early phonological acquisition*. PhD thesis, University of California, Los Angeles.

Lin, Y. and Mielke, J. (2008). Discovering place and manner features: What can be learned from acoustic and articulatory data. *University of Pennsylvania Working Papers in Linguistics*, 14(1):19.

Lisker, L. (1986). "voicing" in english: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and speech*, 29(1):3–11.

Locke, J. L. (1983). *Phonological acquisition and change*. Academic Pr.

Löfqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, 85(3):1314–1321.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training japanese listeners to identify english/r/and/l: A first report. *The Journal of the Acoustical Society of America*, 89(2):874–886.

Longobardi, E., Rossi-Arnaud, C., Spataro, P., Putnick, D. L., and Bornstein, M. H. (2015). Children's acquisition of nouns and verbs in italian: contrasting the roles of frequency and positional salience in maternal language. *Journal of child language*, 42(1):95–121.

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. research on speech perception. *Research on Speech Perception Technical Report*, 6.

Lukatela, G. and Turvey, M. (1991). Phonological access of the lexicon: Evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance*, 17(4):951.

MacKain, K. S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, 9(3):527–542.

Macken, M. A. and Ferguson, C. A. (1983). Cognitive aspects of phonological development: Model, evidence, and issues. *Children's language*, 4:255–282.

Maddieson, I. (1984). The effects on f0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics*, 12(1):9–15.

Mandel, D. R., Jusczyk, P. W., and Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, 6(5):314–317.

Mani, N. and Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, 15(5):445–470.

Mann, V. A. and Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, 69(2):548–558.

Marchman, V. A. and Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, 11(3).

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., and Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, pages i–178.

Markman, E. M. and Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2):121–157.

Markman, E. M., Wasow, J. L., and Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, 47(3):241–275.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.

Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, 10(1):29–63.

Martinet, A. (1952). Function, structure, and sound change. *Word*, 8(1):1–32.

Maye, J. and Gerken, L. (2000). Learning phonemes without minimal pairs. In *Proceedings of the 24th annual Boston university conference on language development*, volume 2, pages 522–533. Citeseer.

Maye, J., Weiss, D. J., and Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental science*, 11(1):122–134.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3):B101–B111.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4):539.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (2005). Representation of lexical form: evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6):1308.

McMurray, B., Aslin, R. N., and Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental science*, 12(3):369–378.

McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6):1113–1126.

Menn, L. (1976). *Pattern, control, and contrast in beginning speech: A case study in the development of word form and word function.* PhD thesis, University of Illinois at Urbana-Champaign.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. *Language production*, 2:3–50.

Menn, L. and Vihman, M. (2011). Features in child phonology. *Where Do Phonological Features Come From?: Cognitive, Physical and Developmental Bases of Distinctive Speech Categories*, 6:261.

Menyuk, P., Menn, L., and Silber, R. (1979). Early strategies for the perception and production of words and sounds. *Language acquisition*, pages 49–70.

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*, 123(2):899–909.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91(1):3.

Mielke, J. (2008). *The emergence of distinctive features.* Oxford University Press.

Mitterer, H. and Ernestus, M. (2006). Listeners recover/t/s that speakers reduce: Evidence from/t/-lenition in dutch. *Journal of Phonetics*, 34(1):73–103.

Mitterer, H. and Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1):168–173.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of japanese and english. *Perception & Psychophysics*, 18(5):331–340.

Morton, J. and Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, 15(1):43–51.

Munson, B., Edwards, J., and Beckman, M. E. (2005a). Phonological knowledge in typical and atypical speech–sound development. *Topics in language disorders*, 25(3):190–206.

Munson, B., Kurtz, B. A., and Windsor, J. (2005b). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 48(5):1033–1047.

Nadeu, M. (2014). Stress-and speech rate-induced vowel quality variation in catalan and spanish. *Journal of Phonetics*, 46:1–22.

Narayan, C. R., Werker, J. F., and Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, 13(3):407–420.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98(1):13–30.

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *The Journal of the Acoustical Society of America*, 110(3):1598–1605.

Nittrouer, S. and Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language, and Hearing Research*, 30(3):319–329.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, 32(1):120–132.

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2):204–238.

Odden, D. (2006). Phonology ex nihilo aka radical substance-free phonology and why i might

recant. https://drive.google.com/file/d/1aZxZmpxYIN-umIdev0PkElX1OU1YGi9G/view. Online; accessed 29 January 2020.

Ohala, J. (1973). Explanations for the intrinsic pitch of vowels. *Monthly Internal Memorandum, Phonology Laboratory, University of California at Berkeley*, pages 9–26.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech*, pages 189–216. Springer.

Ohala, J. J. (1990). There is no interface between phonology and phonetics: a personal view. *Journal of phonetics*, 18(2):153–172.

Ohala, J. J. (1993). The phonetics of sound change. *Historical linguistics: Problems and perspectives*, pages 237–278.

Osthoff, H. and Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, volume 1. S. Hirzel, Leipzig.

Ota, M. (2006). Input frequency and word truncation in child japanese: Structural and lexical effects. *Language and Speech*, 49(2):261–294.

Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6):445–449.

Pater, J., Stager, C., and Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, pages 384–402.

Paul, R. and Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of Speech, Language, and Hearing Research*, 35(1):99–107.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593):604–607.

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., and Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 38(4):616–624.

Peperkamp, S. and Dupoux, E. (2007). Learning the mapping from surface to underlying representations in an artificial language. *Laboratory phonology*, 9:315–338.

Peperkamp, S., Le Calvez, R., Nadal, J.-P., and Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, 101(3):B31–B41.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological studies in language*, 45:137–158.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3):115–154.

Pisoni, D. B. and Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, 55(2):328–333.

Polka, L., Colantonio, C., and Sundara, M. (2001). A cross-language comparison of/d/–/ð/perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, 109(5):2190–2201.

Polka, L. and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human perception and performance*, 20(2):421.

Pollack, I. and Pickett, J. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, 6(3):165–171.

Pye, C., Ingram, D., and List, H. (1987). A comparison of initial consonant acquisition in

english and quiché. *Keith E. Nelson and Ann Van Kleeck, editors, Children's Language*, 6:175–190.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464):349–351.

Reiss, C. (2018). Substance free phonology. *The Routledge handbook of phonological theory*, pages 425–452.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & psychophysics*, 30(3):217–227.

Rescorla, L. and Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (sli-e). *Journal of Speech, Language, and Hearing Research*, 39(1):153–165.

Rubenstein, H., Lewis, S. S., and Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of verbal learning and verbal behavior*, 10(6):645–657.

Saffran, J., Aslin, R., and Newport, E. (1996). Statistical learning by 8-month-old infants. *Science (New York, NY)*, 274(5294):1926–1928.

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5):1124.

Samuels, B. (2011). A minimalist program for phonology. *The Oxford handbook of linguistic Minimalism*, pages 574–94.

Sandler, W. (1993). Sign language and modularity. *Lingua*, 89(4):315–351.

Serniclaes, W., Ventura, P., Morais, J., and Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, 98(2):B35–B44.

Shukla, M., White, K. S., and Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, 108(15):6038–6043.

Singh, L., Morgan, J. L., and White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51(2):173–189.

Singh, L., White, K. S., and Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, 4(2):157–178.

Smith, B. L., McGregor, K. K., and Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics*, 27(3):355–375.

Smith, L. and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3):1558–1568.

Smith, N. V. et al. (1973). *The acquisition of phonology: A case study*. Cambridge University Press.

Sonderegger, M. and Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (CogSci10).[TFJ]*.

Sosa, A. V. and Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language*, 33(1):31–50.

Stager, C. L. and Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640):381–382.

Stoel-Gammon, C. (1991). Normal and disordered phonology in two-year-olds. *Topics in language disorders*.

Stoel-Gammon, C. and Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child language*, 11(2):247–271.

Stokoe, W. (1960). Sign language structure: An outline of the visual communication system of the american deaf. *Studies in Linguistics: Occasional Papers*, 8.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44(6):1321–1337.

Storkel, H. L. and Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, 26(2):191–211.

Storkel, H. L. and Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *clinical linguistics & phonetics*, 14(6):407–425.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental science*, 8(5):432–443.

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, 60(2):252–269.

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental psychology*, 52(7):1011.

Swingley, D. and Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science*, 13(5):480–484.

Taft, M. and Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, 24(3):320–335.

Thiessen, E. D. and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, 39(4):706.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, 19(2):333–363.

Tincoff, R. and Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10(2):172–175.

Tincoff, R. and Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4):432–444.

Toscano, J. C. and McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, 34(3):434–464.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development*, pages 466–472.

Treiman, R., Baron, J., et al. (1981). Segmental analysis ability: Development and relation to reading ability. *Reading research: Advances in theory and practice*, 3:159–198.

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., and Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148:117–135.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., and Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, 104(33):13273–13278.

Van der Feest, S. V. and Fikkert, P. (2015). Building phonological lexical representations. *Phonology*, 32(2):207–239.

Van der Hulst, H. (1993). Units in the analysis of signs. *Phonology*, 10(02):209–241.

Van Orden, G. C. (1987). A rows is a rose: Spelling, sound, and reading. *Memory & cognition*, 15(3):181–198.

Vihman, M. and Miller, R. (1986). Words and babble at the threshold of lexical acquisitions. In Goldstein, L., Whalen, D., and Best, C., editors, *The Emergent Lexicon: The child's development of a linguistic vocabulary*. Academic Press, New York.

Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman*, pages 69–84. Erlbaum Hillsdale, NJ.

Vihman, M. M. (1992). Early syllables and the construction of phonology. *Phonological development: Models, research, implications*, pages 393–422.

Vihman, M. M. (2014). *Phonological development: The first two years*. John Wiley & Sons.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, 108(1):1–27.

Vihman, M. M., DePaolis, R. A., and Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, 64(s2):121–140.

Vihman, M. M., Kay, E., de Boysson-Bardies, B., Durand, C., and Sundberg, U. (1994). External sources of individual differences? a cross-linguistic analysis of the phonetics of mothers' speech to 1-yr-old children. *Developmental Psychology*, 30(5):651.

Vihman, M. M. and Keren-Portnoy, T. (2013). *The emergence of phonology: Whole-word approaches and cross-linguistic evidence*. Cambridge University Press.

Vihman, M. M., Nakai, S., DePaolis, R. A., and Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3):336–353.

Vihman, M. M. and Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*, 57(2-4):255–266.

Vitevitch, M. S. and Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological science*, 9(4):325–329.

Vitevitch, M. S. and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of memory and language*, 40(3):374–408.

Vitevitch, M. S. and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in english. *Behavior Research Methods, Instruments, & Computers*, 36(3):481–487.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental review*, 13(3):286–350.

Walley, A. C., Smith, L. B., and Jusczyk, P. W. (1986). The role of phonemes and syllables in the perceived similarity of speech sounds for children. *Memory & Cognition*, 14(3):220–229.

Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language*, pages 9–25.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training american listeners to perceive mandarin tones. *The Journal of the acoustical society of America*, 106(6):3649–3658.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917):392–393.

Wedel, A., Kaplan, A., and Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186.

Werker, J. F. and Curtin, S. (2005). Primir: A developmental framework of infant speech processing. *Language learning and development*, 1(2):197–234.

Werker, J. F., Fennell, C. T., Corcoran, K. M., and Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1):1–30.

Werker, J. F. and Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental psychology*, 24(5):672.

Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63.

Werker, J. F., Yeung, H. H., and Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, 21(4):221–226.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the english [s]–[š] boundary. *The Journal of the Acoustical Society of America*, 69(1):275–282.

Whalen, D. H., Levitt, A. G., and Wang, Q. (1991). Intonational differences between the reduplicative babbling of french-and english-learning infants. *Journal of Child Language*, 18(3):501–516.

Yang, B. (1996). A comparative study of american english and korean vowels produced by male and female speakers. *Journal of phonetics*, 24(2):245–261.

Yang, C. D. (2000). Internal and external forces in language change. *Language variation and change*, 12(3):231–250.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in cognitive sciences*, 8(10):451–456.

Yeung, H. H. and Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2):234–243.

Yoshida, K. A., Fennell, C. T., Swingley, D., and Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental science*, 12(3):412–418.

Zamuner, T. S., Gerken, L., and Hammond, M. (2005). The acquisition of phonology based on input: A closer look at the relation of cross-linguistic and child language data. *Lingua*, 115(10):1403–1426.

Zsiga, E. C. (1992). Acoustic evidence for gestural overlap in consonant sequences. *Haskins Laboratories Status Report on Speech Research*, 111(112):43–62.